DEX-0281

CE

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
C12N 15/12, C07K 14/47, 16/18, C12N 15/85, 5/10, C12Q 1/68, G01N 33/68, A01K 67/027

A1

(11) International Publication Number: **WO 99/27091**

(43) International Publication Date: 3 June 1999 (03.06.99)

(21) International Application Number: PCT/GB98/03485

(22) International Filing Date: 23 November 1998 (23.11.98)

(30) Priority Data:
9724828.0 21 November 1997 (21.11.97) GB

(71) Applicant *(for all designated States except US)*: THE UNIVERSITY COURT OF THE UNIVERSITY OF GLASGOW [GB/GB]; University Avenue, Glasgow G12 8QQ (GB).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: LA THANGUE, Nicholas, Barrie [GB/GB]; University of Glasgow, Division of Biochemistry and Molecular Biology, Institute of Biomedical and Life Sciences, Davidson Building, Glasgow G12 8QQ (GB). DE LA LUNA, Susana [ES/ES]; Hospital Duran Ei Reynals, Centre Di Genetica Medica Molecular, E–Barcelona (ES).

(74) Agents: BRASNETT, Adrian, H. et al.; Mewburn Ellis, York House, 23 Kingsway, London WC2B 6HP (GB).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: DP TRANSCRIPTION FACTOR–INTERACTING PROTEIN AND ITS USE

(57) Abstract

The present invention relates to a protein, "DIP", which interacts with the DP transcription factor DP–3α. The protein is provided in isolated form, and the invention provides assays for modulators of the cell cycle which target the DIP–DP–3α interaction.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

DP TRANSCRIPTION FACTOR-INTERACTING PROTEIN AND ITS USE

Field of the Invention.

The present invention relates to a novel gene which encodes a product which functions as a transcriptional regulator.

Background to the Invention.

The E2F family of transcription factors plays a critical role in orchestrating early cell cycle progression. In concert with their afferent regulators, which includes the retinoblastoma tumour suppressor protein pRb and G1 cyclin-dependent kinases (cdks), E2F integrates and co-ordinates early cell cycle events with the transcription of genes required for entry into S phase (Nevins, 1992; LaThangue, 1994; Lam and LaThangue, 1994). Significantly, the pathway responsible for regulating E2F is aberrant in most human tumour cells. For example, the Rb gene frequently suffers inactivating mutations and cyclin D, a critical regulator of pRb activity, is expressed at high levels in certain tumour cells, events which underscore the importance of E2F in cell cycle control.

It is known that physiological E2F arises when a member of two families of proteins, E2F and DP, interact as E2F/DP heterodimers. Each E2F component harbours a trans activation domain in the C-terminal region which also physically associates with an appropriate pocket protein, an interaction that impedes transcriptional activation. For example, E2F-1 is principally regulated by pRb, and E2F-4 by p107 and p130. Other levels of control that influence E2F include phosphorylation, ubiquitination and regulated intracellular location, thus providing additional mechanisms that govern the overall activity of E2F.

Each E2F protein requires a DP protein as an obligate heterodimeric partner. DP-1 is a widespread partner for E2F family members in many cell types (Girling *et al*, 1993) in contrast to, for example, DP-3 which appears to be a less frequent component (Rogers *et al*, 1996). In this respect, it is notable that murine DP-3 differs from other members of the

E2F and DP families in that the DP-3 RNA undergoes extensive processing due to alternative splicing (Ormonroyd *et al*, 1995). Processing events in the 5' untranslated and coding region give rise to spliced variants that are restricted in both cells and tissues. Considering the DP-3 proteins, four distinct isoforms have been identified, referred to as α, β, γ and δ; in addition a truncated human alpha isoform has been designated human DP-2 (Wu *et al*, 1995).

Studies on the properties of the DP-3 isoforms has allowed some novel and important regulatory mechanisms in the control of E2F activity to be uncovered. For example, the α and δ isoforms share an alternatively spliced exon, which encodes a sequence of 16 amino acid residues referred to as the E region, and absent from the β and γ isoforms. An analysis of the role of the E region found that it functions as a nuclear localization signal (NLS), specifically in supplying one half of a bi-partite NLS, thus enabling the incumbent DP protein to undergo nuclear accumulation. Importantly, the presence of the E region, and hence nuclear accumulation, endows on DP proteins an ability to promote cell cycle progression when complexed with E2F family members, such as E2F-4 and -5, which lack an intrinsic NLS. In sharp contrast, nuclear accumulation of the E2F heterodimer mediated through pocket protein binding impedes cell cycle progression, thus highlighting the importance of intracellular distribution as a means towards controlling E2F activity.

The BTB/POZ domain is a highly conserved protein-protein interaction domain present in a variety of proteins, many of which have DNA-related functions (Bardwell and Treisman, 1994; Zollman *et al*, 1994; Albagli *et al*, 1995). For example, the Drosophila GAGA transcription factor, which is implicated in regulating nucleosome and chromatin re-organisation, contains a BTB/POZ domain which enables both homo and heterodimeric interactions with other BTB/POZ domains to occur. However, there are some examples of proteins in which the BTB/POZ domain has been shown to inhibit DNA binding. For example, the

BTB/POZ domain in ZID is believed to interfere with the DNA binding activity of the ZID zinc fingers, although in other cases an inhibitory activity of the BTB/POZ domain was not apparent. It is likely therefore that the role of the BTB/POZ domain, apart from as an interface that allows protein-protein interactions, will be protein-dependent.

Moreover, a number of BTB/POZ domain proteins may be important in tumorigenesis. The gene encoding LAZ3/BCL6 is disrupted by chromosomal translocation in certain lymphomas, and similarly in acute myeloid leukaemia PLZF is fused to the retinoic receptor a gene. Although the contribution of these translocation events to the escape from normal growth control remains unclear, such studies nevertheless do imply that some BTB/POZ-domain proteins have a role in regulating proliferation.

Summary of the invention.

We have now defined a new pathway of growth control that converges on E2F. Specifically, certain DP components of the E2F/DP heterodimer physically associate with a novel BTB/POZ domain protein, referred to as DIP (for DP-interacting protein), which possesses significant identify to the product of the Drosophila gene germ cell-less. DIP is a potent transcriptional repressor and can inactivate DP-dependent transcription. In mammalian cells, DIP has a dominant influence on the intracellular distribution of DP proteins by directing them into a characteristic speckled nuclear pattern. Importantly, the expression of DIP causes early cell cycle arrest but, in contrast, when co-expressed with DP-3 in some circumstances can promote apoptosis. Thus, together with the well-documented regulation of E2F by the pRb pathway, our study defines a new pathway of growth control, mediated through DIP, that converges on E2F.

To identify DIP, we used a two-hybrid assay system as described in the accompanying examples. DIP has a predicted molecular

weight of 68 kDa and lacks significant similarity to other proteins apart from the *Drosophila melanogaster* gene *germ cell-less*. The identification of this protein and its gene provides, *inter alia*, novel nucleic acids, polypeptides and related products, and novel assay methods useful in identifying novel regulators of the cell cycle.

The present invention thus provides an isolated polypeptide which comprises residues 1 to 524 of SEQ ID NO. 2 or a polypeptide having at least 70% sequence identity to SEQ ID NO. 2. The invention further provides active portions and fragments which comprises an epitope of said polypeptide. Unless otherwise specified below, such portions and fragments are also referred to as a polypeptide of the invention.

In another aspect, the invention provides an antibody capable of binding a polypeptide of the invention, such as a monoclonal antibody.

The invention also provides an isolated nucleic acid which encodes a polypeptide of the invention, including a nucleic acid which comprises or consists essentially of nucleotides encoding the open reading frame (199 to 1770) of SEQ ID NO:1 or the complement thereof. The invention further provides a nucleic acid capable of selectively hybridizing to either strand of SEQ ID NO:1 such as a nucleic acid which has at least 70% homology to SEQ ID NO:1. Fragments of such selectively hybridizing nucleic acids are also part of the invention. Also provided by the present invention are oligonucleotides which consist essentially of from 15 to 50 contiguous nucleotides of the nucleic acids mentioned above.

Unless specified to the contrary, all the above described nucleic acids are referred to as a "nucleic acid" or a "polynucleotide" of the invention.

The nucleic acids may be in the form of a vector, such as an expression vector wherein said nucleic acid is operably linked

to a promoter heterologous to said nucleic acid. The promoter will be compatible with a desired host cell, and such host cells form a further aspect of the invention.

Nucleic acids encoding or associated with the *DIP* gene may be used in methods of detecting the presence or absence of said gene in a human or non-human mammalian subject, said method comprising;

(a) bringing a sample of nucleic acid from said subject into contact, under hybridizing conditions, with a polynucleotide of the invention; and

(b) determining whether said polynucleotide has been able to hybridize to a homologous sequence in said nucleic acid.

The method may be performed using a polynucleotide primer suitable for use in a polymerase chain reaction (PCR), and the determining may be performed in conjunction with a second primer using PCR such that a portion of the *DIP* gene is amplified.

In some instances, it the determining step may include determining the sequence of the *DIP* gene, when present, in the nucleic acid sample. As one alternative, restriction length fragment polymorphisms associated with the gene may be established and the assay performed with a sample which has been digested with a restriction enzyme. Another method of determining is via PCR length polymorphisms, for example through variation in the sizes of introns. Other specific means of determining hybridization are well known and routine in the art and may also be used.

The invention further provides immunological assays which comprise:

(a) bringing a body sample from said subject into contact, under binding conditions, with an antibody of the invention; and

(b) determining whether said antibody has been able to bind to a polypeptide in said sample.

The invention further provides assays for modulators of cell cycle progression which assays are based on the finding of the interaction of DIP with DP proteins, as well as other proteins which comprise a BTB/POZ domain. These include interactions with other DIP family members and interactions which provide DIP homodimers. Assays may take any suitable format, as described in detail below.

As used herein, "comprise(s)" and "comprising" are to be interpreted as "include(s) and "including".

The percentage homology (also referred to as identity) of DNA and amino acid sequences can be calculated using commercially available algorithms. The following programs (provided by the National Center for Biotechnology Information) may be used to determine homologies: BLAST, gapped BLAST, BLASTN and PSI-BLAST, which may be used with default parameters.

The algorithm GAP (Genetics Computer Group, Madison, WI). GAP uses the Needleman and Wunsch algorithm to align two complete sequences that maximizes the number of matches and minimizes the number of gaps. Generally, the default parameters are used, with a gap creation penalty = 12 and gap extension penalty = 4. Use of either of the terms "homology" and "homologous" herein does not imply any necessary evolutionary relationship between compared sequences, in keeping for example with standard use of terms such as "homologous recombination" which merely requires that two nucleotide sequences are sufficiently similar to recombine under the appropriate conditions.

Another method for determining the best overall match between a nucleic acid sequence or a portion thereof, and a query sequence is the use of the FASTDB computer program based on the algorithm of Brutlag *et al* (Comp. App. Biosci., 6; 237-245

(1990)). The program provides a global sequence alignment. The result of said global sequence alignment is in percent identity. Suitable parameters used in a FASTDB search of a DNA sequence to calculate percent identity are: Matrix=Unitary, k-tuple=4, Mismatch penalty=1, Joining Penalty=30, Randomization Group Length=0, Cutoff Score=1, Gap Penalty=5, Gap Size Penalty=0.05, and Window Size=500 or query sequence length in nucleotide bases, whichever is shorter. Suitable parameters to calculate percent identity and similarity of an amino acid alignment are: Matrix=PAM 150, k-tuple=2, Mismatch Penalty=1, Joining Penalty=20, Randomization Group Length=0, Cutoff Score=1, Gap Penalty=5, Gap Size Penalty=0.05, and Window Size=500 or query sequence length in nucleotide bases, whichever is shorter.

Detailed description of the invention.

A. Polypeptides

Isolated polypeptides of the invention will be those as defined above in isolated form, free or substantially free of material with which it is naturally associated such as other polypeptides with which it is found in the cell. The polypeptides may of course be formulated with diluents or adjuvants and still for practical purposes be isolated - for example the polypeptides will normally be mixed with gelatin or other carriers if used to coat microtitre plates for use in immunoassays. The polypeptides may be glycosylated, either naturally or by systems of heterologous eukaryotic cells, or they may be (for example if produced by expression in a prokaryotic cell) unglycosylated. Polypeptides may phosphorylated and/or acetylated.

A polypeptide of the invention may also be in a substantially purified form, in which case it will generally comprise the polypeptide in a preparation in which more than 90%, e.g. 95%, 98% or 99% of the polypeptide in the preparation is a polypeptide of the invention.

Polypeptides of the invention may be modified for example by the addition of histidine residues to assist their purification or by the addition of a signal sequence to promote their secretion from a cell.

Polypeptides which are amino acid sequence variants, alleles, derivatives or mutants are also provided by the present invention, such forms having at least 70% sequence identity, for example at least 80%, 90%, 95%, 98% or 99% sequence identity to SEQ ID NO. 2. A polypeptide which is a variant, allele, derivative or mutant may have an amino acid sequence which differs from that given in SEQ ID NO. 2 by one or more of addition, substitution, deletion and insertion of one or more (such as from 1 to 20, for example 2, 3, 4, or 5 to 10) amino acids.

Preferred such polypeptides include those which are encoded by the *DIP* gene of other mammals, particularly primates and most particularly man, as well as fragments of such polypeptides, such fragments being those as defined above. The primary sequence of the DIP protein will be substantially similar to that of SEQ ID NO:2 and may be determined by routine techniques available to those of skill in the art. In essence, such techniques comprise using polynucleotides of the present invention as probes to recover and to determine the sequence of the *DIP* gene in other species. A wide variety of techniques are available for this, for example PCR amplification and cloning of the gene using a suitable source of mRNA (e.g. from an embryo or an actively dividing differentiated or tumour cell), or by methods comprising obtaining a cDNA library from the mammal, e.g a cDNA library from one of the above-mentioned sources, probing said library with a polynucleotide of the invention under stringent conditions, and recovering a cDNA encoding all or part of the DIP protein of that mammal. Where a partial cDNA is obtained, the full length coding sequence may be determined by primer extension techniques.

An "active portion" of the polypeptides means a peptide which is less than said full length polypeptide, but which retains its essential biological activity. In particular, the active portion retains the ability to interact with DP-3α in mammalian or yeast cells. Suitable active portions thus include the portion of SEQ ID NO:2 between about residues 105 and 198, for example between 100 to 300, 100 to 250, 50 to 300 and 50 to 250, as well as variants of such segments which retain the ability to interact with DP-3α.

Active portions may also include those which are phosphorylated and/or acetylated, particularly in a cell-cycle specific manner.

Active portions may be used in methods of therapy including gene therapy.

An "inactive portion" of the polypeptides means a peptide which is still identifiable as a polypeptide of the invention but which through mutation is truncated or internally deleted. Examples of such polypeptides will include those which comprise at least 20, for example at least 30, 40, 50, 75 or 100 contiguous amino acids derived from SEQ ID NO:2, its variants including its species homologues.

Inactive portions may include at least one epitope to which antibodies are able to bind specifically.

Inactive portions may include fragments of the above-mentioned active portions which are capable of competing with the full length human or murine DIP protein for binding to a DP protein such as DP-3α. Preferably such fragments are those which are capable of antagonizing the formation of a DIP-DP-3α complex or DIP homodimerization, under conditions suitable for such complex formation or homodimerization to take place in the absence of such an inactive portion. Inactive portions also include dominant negative mutants of DIP.

A "fragment" means a stretch of amino acid residues of at least about five to seven contiguous amino acids, often at least about seven to nine contiguous amino acids, typically at least about nine to 13 contiguous amino acids and, most preferably, at least about 20 to 30 or more contiguous amino acids. Fragments of the polypeptides comprise epitopes useful for raising antibodies to a portion of the amino acid sequences of SEQ ID NO. 2 or variants with at least 70% homology to SEQ ID NO. 2. Preferred epitopes are those to which antibodies are able to bind specifically, as defined below in section B.

As defined above, some fragments of the invention may be either active or inactive portions.

A polypeptide according to the present invention may be isolated and/or purified (e.g. using an antibody) for instance after production by expression from encoding nucleic acid (for which see below). Polypeptides according to the present invention may also be generated wholly or partly by chemical synthesis, for example in a step-wise manner. The isolated and/or purified polypeptide may be used in formulation of a composition, which may include at least one additional component, for example a pharmaceutical composition including a pharmaceutically acceptable excipient, vehicle or carrier. A composition including a polypeptide according to the invention may be used in prophylactic and/or therapeutic treatment as discussed below.

A polypeptide according to the present invention may be used as an immunogen or otherwise in obtaining specific antibodies. Antibodies are useful in purification and other manipulation of polypeptides and peptides, diagnostic screening and therapeutic contexts. This is discussed further below.

A polypeptide according to the present invention may be used in screening for molecules which affect or modulate its activity or function. Such molecules may be useful in a therapeutic (possibly including prophylactic) context, for example in

connection with conditions which involve abnormal or aberrant expression of DIP.

A polypeptide of the invention may be labelled with a revealing label. The revealing label may be any suitable label which allows the polypeptide to be detected. Suitable labels include radioisotopes, e.g. ^{121}I, enzymes, antibodies, polynucleotides and linkers such as biotin. Labelled polypeptides of the invention may be used in diagnostic procedures such as immunoassays in order to determine the amount of a polypeptide of the invention in a sample. Polypeptides or labelled polypeptides of the invention may also be used in serological or cell mediated immune assays for the detection of immune reactivity to said polypeptides in animals and humans using standard protocols.

A polypeptide or labelled polypeptide of the invention or fragment thereof may also be fixed to a solid phase, for example the surface of an immunoassay well or dipstick.

Such labelled and/or immobilized polypeptides may be packaged into kits in a suitable container along with suitable reagents, controls, instructions and the like.

Such polypeptides and kits may be used in methods of detection of antibodies to such polypeptides present in a sample or active portions or fragments thereof by immunoassay.

Immunoassay methods are well known in the art and will generally comprise:

(a) providing a polypeptide comprising an epitope bindable by an antibody against said protein;

(b) incubating a biological sample with said polypeptide under conditions which allow for the formation of an antibody-antigen complex; and

(c) determining whether antibody-antigen complex comprising said polypeptide is formed.

The identification of the polypeptide expressed by the *DIP* gene enables assays to be developed to identify further cellular proteins with which the polypeptide is associated, in addition to DP-3. For example, polypeptides of the present invention may be required in a regulatory pathway in which their function is to interact with other factors which in turn promote or maintain essential cellular functions associated with cell cycle control. The polypeptides of the present invention may be used in two-hybrid assays to determine cellular factors with which they become associated.

Two-hybrid assays may be in accordance with those disclosed by Fields and Song, 1989, Nature 340; 245-246. In such an assay the DNA binding domain (DBD) and the transcriptional activation domain (TAD) of the yeast GAL4 transcription factor are fused to the first and second molecules respectively whose interaction is to be investigated. A functional GAL4 transcription factor is restored only when two molecules of interest interact. Thus, interaction of the molecules may be measured by the use of a reporter gene operably linked to a GAL4 DNA binding site which is capable of activating transcription of said reporter gene. Other transcriptional activator domains may be used in place of the GAL4 TAD, for example the viral VP16 activation domain. In general, fusion proteins comprising DNA binding domains and activation domains may be made.

In the present case polypeptides of the invention may be expressed as fusion proteins with an appropriate domain and candidate second polypeptides with which those of the invention might associate can be produced as fusion proteins with an appropriate corresponding domain. Alternatively libraries such as phage display libraries of such fusion proteins may be screed with a fusion polypeptide of the invention.

B. Antibodies

The provision of the novel polypeptides enables for the first time the production of antibodies able to bind it specifically. Such an antibody may be specific in the sense of being able to distinguish between the polypeptide it is able to bind and other polypeptides of the same species for which it has no or substantially no binding affinity (e.g. a binding affinity of at least about 1000x worse). Specific antibodies bind an epitope on the molecule which is either not present or is not accessible on other molecules. Antibodies according to the present invention may be specific for the wild-type polypeptide. Antibodies according to the invention may be specific for a particular mutant, variant, allele or derivative polypeptide as between that molecule and the wild-type polypeptide, so as to be useful in diagnostic and prognostic methods as discussed below. Antibodies are also useful in purifying the polypeptide or polypeptides to which they bind, e.g. following production by recombinant expression from encoding nucleic acid.

Preferred antibodies according to the invention are isolated, in the sense of being free from contaminants such as antibodies able to bind other polypeptides and/or free of serum components. Monoclonal antibodies are preferred for some purposes, though polyclonal antibodies are within the scope of the present invention.

Antibodies may be obtained using techniques which are standard in the art. Methods of producing antibodies include immunising a mammal (e.g. mouse, rat, rabbit) with a polypeptide of the invention. Antibodies may be obtained from immunised animals using any of a variety of techniques known in the art, and screened, preferably using binding of antibody to antigen of interest. For instance, Western blotting techniques or immunoprecipitation may be used (Armitage *et al*, Nature, 357:80-82, 1992).

As an alternative or supplement to immunising a mammal with a peptide, an antibody specific for a protein may be obtained

from a recombinantly produced library of expressed immunoglobulin variable domains, e.g. using lambda bacteriophage or filamentous bacteriophage which display functional immunoglobulin binding domains on their surfaces; for instance see WO92/01047.

Antibodies according to the present invention may be modified in a number of ways. Indeed the term "antibody" should be construed as covering any binding substance having a binding domain with the required specificity. Thus the invention covers antibody fragments, derivatives, functional equivalents and homologues of antibodies, including synthetic molecules and molecules whose shape mimics that of an antibody enabling it to bind an antigen or epitope.

Example antibody fragments, capable of binding an antigen or other binding partner are the Fab fragment consisting of the VL, VH, Cl and CH1 domains; the Fd fragment consisting of the VH and CH1 domains; the Fv fragment consisting of the VL and VH domains of a single arm of an antibody; the dAb fragment which consists of a VH domain; isolated CDR regions and F(ab')2 fragments, a bivalent fragment including two Fab fragments linked by a disulphide bridge at the hinge region. Single chain Fv fragments are also included.

Humanized antibodies in which CDRs from a non-human source are grafted onto human framework regions, typically with the alteration of some of the framework amino acid residues, to provide antibodies which are less immunogenic than the parent non-human antibodies, are also included within the present invention

A hybridoma producing a monoclonal antibody according to the present invention may be subject to genetic mutation or other changes. It will further be understood by those skilled in the art that a monoclonal antibody can be subjected to the techniques of recombinant DNA technology to produce other antibodies or chimeric molecules which retain the specificity

of the original antibody. Such techniques may involve introducing DNA encoding the immunoglobulin variable region, or the complementarity determining regions (CDRs), of an antibody to the constant regions, or constant regions plus framework regions, of a different immunoglobulin. See, for instance, EP-A-184187, GB-A-2188638 or EP-A-0239400. Cloning and expression of chimeric antibodies are described in EP-A-0120694 and EP-A-0125023.

Hybridomas capable of producing antibody with desired binding characteristics are within the scope of the present invention, as are host cells, eukaryotic or prokaryotic, containing nucleic acid encoding antibodies (including antibody fragments) and capable of their expression. The invention also provides methods of production of the antibodies including growing a cell capable of producing the antibody under conditions in which the antibody is produced, and preferably secreted.

The reactivities of antibodies on a sample may be determined by any appropriate means. Tagging with individual reporter molecules is one possibility. The reporter molecules may directly or indirectly generate detectable, and preferably measurable, signals. The linkage of reporter molecules may be directly or indirectly, covalently, e.g. via a peptide bond or non-covalently. Linkage via a peptide bond may be as a result of recombinant expression of a gene fusion encoding antibody and reporter molecule.

One favoured mode is by covalent linkage of each antibody with an individual fluorochrome, phosphor or laser dye with spectrally isolated absorption or emission characteristics. Suitable fluorochromes include fluorescein, rhodamine, phycoerythrin and Texas Red. Suitable chromogenic dyes include diaminobenzidine.

Other reporters include macromolecular colloidal particles or particulate material such as latex beads that are coloured, magnetic or paramagnetic, and biologically or chemically active

agents that can directly or indirectly cause detectable signals to be visually observed, electronically detected or otherwise recorded. These molecules may be enzymes which catalyse reactions that develop or change colours or cause changes in electrical properties, for example. They may be molecularly excitable, such that electronic transitions between energy states result in characteristic spectral absorptions or emissions. They may include chemical entities used in conjunction with biosensors. Biotin/avidin or biotin/streptavidin and alkaline phosphatase detection systems may be employed.

The mode of determining binding is not a feature of the present invention and those skilled in the art are able to choose a suitable mode according to their preference and general knowledge.

Antibodies according to the present invention may be used in screening for the presence of a polypeptide, for example in a test sample containing cells or cell lysate as discussed, and may be used in purifying and/or isolating a polypeptide according to the present invention, for instance following production of the polypeptide by expression from encoding nucleic acid therefor. Antibodies may modulate the activity of the polypeptide to which they bind and so, if that polypeptide has a deleterious effect in an individual, may be useful in a therapeutic context (which may include prophylaxis).

Diagnostic and therapeutic uses of antibodies include uses associated with the treatment and diagnosis of conditions associated with germ cell abnormalities, including teratomas and seminomas, where DIP levels and the interaction of DIP play a role in maintaining the cells in an abnormal state.

An antibody may be provided in a kit, which may include instructions for use of the antibody, e.g. in determining the presence of a particular substance in a test sample. One or more other reagents may be included, such as labelling

molecules, buffer solutions, elutants and so on. Reagents may be provided within containers which protect them from the external environment, such as a sealed vial.

C. Nucleic acid.

The "DIP locus" includes the *DIP* gene, both the coding sequence (exons) and intervening sequences (introns), and its regulatory elements for controlling transcription and/or translation.

The term "*DIP* gene" includes normal alleles of the gene which encodes in wild-type mice an mRNA which comprises a sequence substantially corresponding to that of SEQ ID NO. 1. It also includes alleles of this gene carrying one or more variations. The term also includes mammalian species homologues, particularly human homologues.

Our data indicate that the DIP gene product is a transcriptional repressor protein. Proteins of this type are well-conserved between species and those of skill in the art would recognise that a sequence of this nature from a single species is representative of the genus of eukaryotic, particularly invertebrate and vertebrate, more particularly vertebrate and especially mammalian DIP homologues. In turn, the nucleic acid coding sequences for such DIP proteins will be conserved.

Nucleic acid includes DNA (including both genomic and cDNA) and RNA, and also synthetic nucleic acids, such as those with modified backbone structures intended to improve stability of the nucleic acid in a cell. A number of different types of modification to oligonucleotides are known in the art. These include methylphosphonate and phosphorothioate backbones, addition of acridine or polylysine chains at the 3' and/or 5' ends of the molecule. For the purposes of the present invention, it is to be understood that the polynucleotides described herein may be modified by any method available in the art. Such modifications may be carried out in order to enhance

the *in vivo* activity or lifespan of polynucleotides of the invention. Where nucleic acid according to the invention includes RNA, reference to the sequences shown in the accompanying listings should be construed as reference to the RNA equivalent, with U substituted for T.

Nucleic acid of the invention may be single or double stranded polynucleotides. Single stranded nucleic acids of the invention include anti-sense nucleic acids.

The invention further provides ribozymes which comprise a nucleic acid sequence of the invention.

Generally, nucleic acid according to the present invention is provided as an isolate, in isolated and/or purified form, or free or substantially free of material with which it is naturally associated, such as free or substantially free of nucleic acid flanking the gene in the human genome, except possibly one or more regulatory sequence(s) for expression. Nucleic acid may be wholly or partially synthetic and may include genomic DNA, cDNA or RNA.

Nucleic acids of the invention include nucleic acids which comprise a sequence encoding a polypeptide which is selected from the group consisting of residues 1 to 524 of SEQ ID NO. 2 or fragment thereof and a polypeptide having at least 70% sequence identity to SEQ ID NO. 2 or fragment thereof. Preferably the degree of sequence identity in either case is at least 80%, such as at least 90%, 95%, 98% or 99%.

Nucleic acids of the invention further include nucleic acids which comprise a sequence having at least 70% homology, more preferably at least 80%, such as at least 90%, 95%, 98% or 99% sequence homology to the nucleic acid sequences of SEQ ID NO. 1 or its complement.

The invention also provides nucleic acids which are fragments of the nucleic acids described in the two preceding paragraphs. Particular nucleic acids which are preferred include:

(a) nucleic acids which comprise a sequence encoding an active portion of the invention;

(b) nucleic acid fragments of a sequence having at least 70% homology to the nucleic acid sequences of SEQ ID NO. 1 or its complement, said fragments comprising at least 15 nucleotides; and

(c) nucleic acids which consist essentially of from 15 to 50, for example from 15 to 35, 18 to 35, 15 to 24, 18 to 30, 18 to 21 or 21 to 24 nucleotides of a sequence having at least 70% homology to the nucleic acid sequence of SEQ ID NO. 1 or its complement.

The nucleic acids (a), (b) and (c) above are not mutually exclusive. Nucleic acids of categories (a) and (b) will include nucleic acids which comprise at least 15, such as at least 20, 30, 50 or 100 nucleotides.

The term "consist essentially of" refers to nucleic acids which do not include any additional 5' or 3' nucleic acid sequences. In a further aspect of the invention, nucleic acids of the invention which consist essentially of from 15 to 30 nucleotides as defined above may however be linked at the 3' but preferably 5' end to short (e.g from 4 to 15, such as from 4 to 10 nucleotides) additional sequences to which they are not naturally linked. Such additional sequences are preferably linkers which comprise a restriction enzyme recognition site to facilitate cloning when the nucleic acid of the invention is used for example as a PCR primer.

Nucleic acids of the invention, particularly short (less than 50) sequences useful as probes and primers may carry a revealing label. Suitable labels include radioisotopes such as ^{32}P or ^{35}S, fluorescent labels, enzyme labels, or other protein labels such as biotin. Such labels may be added to

polynucleotides or primers of the invention and may be detected using by techniques known per se.

Also included within the scope of the invention are antisense sequences based on the nucleic acid sequences described herein, preferably in the form of oligonucleotides, particularly stabilized oligonucleotides, or ribozymes. Antisense oligonucleotides may be designed to hybridise to the complementary sequence of nucleic acid, pre-mRNA or mature mRNA, interfering with the production of polypeptide encoded by a given DNA sequence (e.g. either native DIP polypeptide or a mutant form thereof), so that its expression is reduce or prevented altogether. Ribozymes will be designed to cleave mRNA encoded by a *DIP* nucleic acid sequence of the invention, desirably at a target sequence specific to the *DIP* sequence. In addition to the coding sequence, antisense techniques can be used to target the control sequences of the *DIP* gene, e.g. in the 5' flanking sequence. The construction of antisense sequences and their use is described in Peyman and Ulman, Chemical Reviews, 90:543-584, (1990), Crooke, Ann. Rev. Pharmacol. Toxicol., 32:329-376, (1992), and Zamecnik and Stephenson, P.N.A.S, 75:280-284, (1974). The construction of ribozymes and their use in described in for instance Gibson and Shillitoe, Molecular Biotechnology 7(2): 125-137, (1997).

Nucleic acid sequences encoding all or part of the *DIP* gene and/or its regulatory elements can be readily prepared by the skilled person using the information and references contained herein and techniques known in the art (for example, see Sambrook, Fritsch and Maniatis, "Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Laboratory Press, 1989, and Ausubel *et al*, Short Protocols in Molecular Biology, John Wiley and Sons, 1992). These techniques include (i) the use of the polymerase chain reaction (PCR) to amplify samples of such nucleic acid, e.g. from genomic sources, (ii) chemical synthesis, or (iii) preparing cDNA sequences. Modifications to the wild type sequences described herein can be made, e.g. using site directed mutagenesis, to lead to the expression of

modified polypeptides or to take account of codon preference in the host cells used to express the nucleic acid.

In general, short sequences for use as primers will be produced by synthetic means, involving a step wise manufacture of the desired nucleic acid sequence one nucleotide at a time. Techniques for accomplishing this using automated techniques are readily available in the art.

Longer polynucleotides (including those from subjects expressing inactive portions) will generally be produced using recombinant means, for example using a PCR (polymerase chain reaction) cloning techniques. This will involve making a pair of primers (e.g. of about 15-50 nucleotides) based on the sequence information provided herein to a region of the mRNA or genomic sequence encoding the mRNA which it is desired to clone, bringing the primers into contact with mRNA or cDNA obtained from a murine or human cell (e.g. a brain cell, particularly a fetal brain cell), performing a polymerase chain reaction under conditions which bring about amplification of the desired region, isolating the amplified fragment (e.g. by purifying the reaction mixture on an agarose gel) and recovering the amplified DNA. The primers may be designed to contain suitable restriction enzyme recognition sites so that the amplified DNA can be cloned into a suitable cloning vector.

Such techniques may be used to obtain all or part of the sequences described herein. Genomic clones containing the *DIP* gene and its introns and promoter regions may also be obtained in an analogous manner, starting with genomic DNA from a murine or human cell, e.g. a primary cell such as a liver cell, a tissue culture cell or a library such as a phage, cosmid, YAC (yeast artificial chromosome), BAC (bacterial artificial chromosome) or PAC (P1/P2 phage artificial chromosome) library.

Polynucleotides which are not 100% homologous to the sequences of the present invention but fall within the scope of the invention can be obtained in a number of ways.

Other murine variants (for example allelic forms) of the *DIP* gene described herein may be obtained for example by probing cDNA or genomic DNA libraries made from murine tissue.

In addition, other animal, for example fish (such as the Zebra fish), worm (such as *C.elegans*) and particularly mammalian (e.g. rat or rabbit, sheep, goat, pig, or primate particularly human) homologues of the *DIP* gene may be obtained. Such sequences may be obtained by making or obtaining cDNA libraries made from dividing cells or tissues or genomic DNA libraries from other animal species, and probing such libraries with probes comprising all or part of a nucleic acid of the invention under conditions of medium to high stringency (for example for hybridization on a solid support (filter) overnight incubation at 42°C in a solution containing 50% formamide, 5xSSC (750 mM NaCl, 75 mM sodium citrate), 50 mM sodium phosphate (pH7.6), 5x Denhardt's solution, 10% dextran sulphate and 20 μg/ml salmon sperm DNA, followed by washing in 0.03M sodium chloride and 0.03M sodium citrate (i.e. 0.2x SSC) at from about 50°C to about 60°C).

Alternatively, such polynucleotides may be obtained by site directed mutagenesis of the sequences of SEQ ID NO. 1 or allelic variants thereof. This may be useful where for example silent codon changes are required to sequences to optimise codon preferences for a particular host cell in which the polynucleotide sequences are being expressed. Other sequence changes may be desired in order to introduce restriction enzyme recognition sites, or to alter the property or function of the polypeptides encoded by the polynucleotides. Further changes may be desirable to represent particular coding changes which are required to provide, for example, conservative substitutions.

In the context of cloning, it may be necessary for one or more gene fragments to be ligated to generate a full-length coding sequence. Also, where a full-length encoding nucleic acid molecule has not been obtained, a smaller molecule representing

part of the full molecule, may be used to obtain full-length clones. Inserts may be prepared from partial cDNA clones and used to screen cDNA libraries. The full-length clones isolated may be subcloned into expression vectors and activity assayed by transfection into suitable host cells, e.g. with a reporter plasmid.

The present invention also extends to nucleic acid comprising transcriptional control sequences for the *DIP* gene. Such control sequences will be found 5' to the open reading frame of the *DIP* gene and are obtainable by probing a genomic DNA library (such as a phage, cosmid, YAC, BAC or PAC library) of a mammal with a nucleic acid of the invention, selecting a clone which hybridizes under conditions of medium to high stringency, and sequencing the clone 5' to the open reading frame of the gene. Where only a small amount of sequence is present in the 5' region, this sequence may be used to reprobe the library to genome walk further upstream. Analysis of the upstream region will reveal control regions for gene expression including control regions common to many genes (i.e TATA and CAAT boxes) and other control regions, usually located from 1 to 10,000, such as 1 to 1000 or 50 to 500 nucleotides upstream of the start of transcription.

To confirm that such regions are the control regions of the gene, they may be linked to a reported gene (such as β-galactosidase) and tested in any suitable *in vitro* or *in vivo* system. For example the construct of the control region (e.g. comprising 50 to 500 nucleotides upstream of the start of transcription) and the reporter gene may be used to produce a transgenic animal (see below) and the pattern of expression, both spatially and developmentally, may be compared with that of the *DIP* gene. Where substantially similar patterns of expression are found, this shows that the construct comprises substantially all of the control region of the wild type gene.

The control region may be mutated to identify specific subregions responsible for transcriptional control. This may

be achieved by a number of techniques well known in the art as such, including DNase protection footprint assays, in which the control region is brought into contact with an extract from a cell in which the *DIP* gene is actively expressed, and the regions of the control region which bind factors in that extract is determined.

Isolated nucleic acid comprising such control regions obtainable by such a method form a further aspect of the present invention.

The present invention further extends to genomic DNA exon sequences found between the introns encoding a *DIP* gene in an animal subject, such as those mentioned above and including humans. Such exon sequences may be obtained in a manner analogous to that described above for the transcriptional control sequences, with the appropriate genome walking being conducted between the intron sequences. The locations of the exons may be determined by comparing genomic and cDNA sequences of the *DIP* gene, observing where the sequences line up and diverge, and looking for consensus splice sequences which define intron/exon boundaries.

Exon sequences obtainable by these or analogous methods may be used in the construction of mini-gene sequences which comprise nucleic acid encoding polypeptides of the invention which sequences are interrupted by one or more exon sequences.

Mini-genes may also be constructed using heterologous exons, derived from any eukaryotic source.

Nucleic acid according to the present invention, such as a full-length coding sequence or oligonucleotide probe or primer, may be provided as part of a kit, e.g. in a suitable container such as a vial in which the contents are protected from the external environment. The kit may include instructions for use of the nucleic acid, e.g. in PCR and/or a method for determining the presence of nucleic acid of interest in a test

sample. A kit wherein the nucleic acid is intended for use in PCR may include one or more other reagents required for the reaction, such as polymerase, nucleosides, buffer solution etc. The nucleic acid may be labelled. A kit for use in determining the presence or absence of nucleic acid of interest may include one or more articles and/or reagents for performance of the method, such as means for providing the test sample itself, e.g. a swab for removing cells from the buccal cavity or a syringe for removing a blood sample (such components generally being sterile). In a further aspect, the present invention provides an apparatus for screening nucleic acid, the apparatus comprising storage means including the a nucleic acid or the invention or fragment thereof, the stored sequence being used to compare the sequence of the test nucleic acid to determine the presence of mutations.

Polynucleotides or primers of the invention or fragments thereof labelled or unlabelled may be used by a person skilled in the art in nucleic acid-based tests for detecting the *DIP* gene in the human or animal body. In the case of detecting, this may be qualitative and/or quantitative. Detection includes analytical steps such as those which involve sequencing the gene in full or in part.

Such tests for detecting generally comprise bringing a human or animal body sample containing DNA or RNA into contact with a probe comprising a polynucleotide or primer of the invention under hybridizing conditions and detecting any duplex formed between the probe and nucleic acid in the sample. Such detection may be achieved using techniques such as PCR or by immobilizing the probe on a solid support, removing nucleic acid in the sample which is not hybridized to the probe, and then detecting nucleic acid which has hybridized to the probe. Alternatively, the sample nucleic acid may be immobilized on a solid support, and the amount of probe bound to such a support can be detected. Suitable assay methods of this any other formats can be found in for example WO89/03891 and WO90/13667.

In one embodiment, the sample nucleic acid may be in the form of whole chromosomes, for example as a metaphase spread. The nucleic acid probe or primer of the invention may be labelled with a fluorescent label to detect the chromosomal location of a *DIP* gene in the spread.

Where a PCR based assay is used in the detection of nucleic acid according to the invention, it is preferred that both primers are nucleic acids according to the present invention. However, one of the two primers may be directed to sequences which flank sequences of the *DIP* gene or are contained within its exons. Those of skill in the art will be able to select specific pairs of PCR primers using routine skill and knowledge in the light of the present disclosure.

A further method of detection according to the invention is in detecting changes to wild-type *DIP* genes, including single base changes, using single stranded conformational polymorphism (SSCP) analysis. Nucleic acid sequence from all or part of a *DIP* DNA or mRNA in a sample is hybridized to a reference sequence, and the mobility of the hybrid is observed in a gel under conditions where any non-hybridized regions within the duplex give rise to changes in mobility.

Nucleic acids of the invention are thus useful in screening a test sample containing nucleic acid for the presence of alleles, mutants and variants, the probes hybridising with a target sequence from a sample obtained from the individual being tested. The conditions of the hybridisation can be controlled to minimise non-specific binding, and preferably stringent to moderately stringent hybridisation conditions are preferred. The skilled person is readily able to design such probes, label them and devise suitable conditions for the hybridisation reactions, assisted by textbooks such as Sambrook *et al* (1989) and Ausubel *et al* (1992).

As well as determining the presence of polymorphisms or mutations in the *DIP* sequence, the probes may also be used to

determine whether mRNA encoding the *DIP* gene is present in a cell or tissue. Such probes are thus useful in the diagnosis of conditions associated with disorders of germ cells, including proliferative disorders such as seminomas and teratomas, and thus the invention provides methods for the diagnosis of such conditions which comprises providing a nucleic acid of the invention and bringing it into contact with a sample of germ cell nucleic acid (or germ cell tissue containing germ cell nucleic acid (in the form of RNA, mRNA or DNA)) under hybridizing conditions and observing the hybridization between said probe and said sample.

Nucleic acid of the invention may be provided in the form of compositions, for example a pharmaceutical composition. Such compositions will include pharmaceutically acceptable carriers and adjuvants. Examples of a suitable carrier include liposomes. Liposomes carrying nucleic acid of the invention (particularly where such nucleic acid is carried by a vector, see below) may be used in methods of gene delivery in gene therapy. Suitable liposome compositions and delivery systems are described in Gill *et al*, Gene Therapy, Vol.4, No.3, pp.199-209 (1997).

D. Vectors.

Nucleic acid polynucleotides of the invention can be incorporated into a recombinant replicable vector. The vector may be used to replicate the nucleic acid in a compatible host cell. Thus in a further embodiment, the invention provides a method of making polynucleotides of the invention by introducing a polynucleotide of the invention into a replicable vector, introducing the vector into a compatible host cell, and growing the host cell under conditions which bring about replication of the vector. The vector may be recovered from the host cell. Suitable host cells are described below in connection with expression vectors.

E. Expression Vectors.

Preferably, a polynucleotide of the invention in a vector is operably linked to a control sequence which is capable of providing for the expression of the coding sequence by the host cell, i.e. the vector is an expression vector.

The term "operably linked" refers to a juxtaposition wherein the components described are in a relationship permitting them to function in their intended manner. A control sequence "operably linked" to a coding sequence is ligated in such a way that expression of the coding sequence is achieved under condition compatible with the control sequences.

Suitable vectors can be chosen or constructed, containing appropriate regulatory sequences, including promoter sequences, terminator fragments, polyadenylation sequences, enhancer sequences, marker genes and other sequences as appropriate. Vectors may be plasmids, viral e.g. 'phage phagemid or baculoviral, cosmids, YACs, BACs, or PACs as appropriate. Vectors include gene therapy vectors, for example vectors based on adenovirus, adeno-associated virus, retrovirus (such as HIV or MLV) or alpha virus vectors.

The vectors may be provided with an origin of replication, optionally a promoter for the expression of the said polynucleotide and optionally a regulator of the promoter. The vectors may contain one or more selectable marker genes, for example an ampicillin resistance gene in the case of a bacterial plasmid or a neomycin resistance gene for a mammalian vector. Vectors may be used *in vitro*, for example for the production of RNA or used to transfect or transform a host cell. The vector may also be adapted to be used *in vivo*, for example in methods of gene therapy. Systems for cloning and expression of a polypeptide in a variety of different host cells are well known. Suitable host cells include bacteria, eukaryotic cells such as mammalian and yeast, and baculovirus systems. Mammalian cell lines available in the art for expression of a heterologous polypeptide include Chinese

hamster ovary cells, HeLa cells, baby hamster kidney cells, COS cells and many others.

Promoters and other expression regulation signals may be selected to be compatible with the host cell for which the expression vector is designed. For example, yeast promoters include S. cerevisiae GAL4 and ADH promoters, S. pombe nmt1 and adh promoter. Mammalian promoters include the metallothionein promoter which is induced in response to heavy metals such as cadmium. Viral promoters such as the SV40 large T antigen promoter or adenovirus promoters may also be used. All these promoters are readily available in the art.

The vectors may include other sequences such as promoters or enhancers to drive the expression of the inserted nucleic acid, nucleic acid sequences so that the polypeptide is produced as a fusion and/or nucleic acid encoding secretion signals so that the polypeptide produced in the host cell is secreted from the cell.

Vectors for production of polypeptides of the invention of for use in gene therapy include vectors which carry a mini-gene sequence of the invention.

Vectors may be transformed into a suitable host cell as described above to provide for expression of a polypeptide of the invention. Thus, in a further aspect the invention provides a process for preparing polypeptides according to the invention which comprises cultivating a host cell transformed or transfected with an expression vector as described above under conditions to provide for expression by the vector of a coding sequence encoding the polypeptides, and recovering the expressed polypeptides. Polypeptides may also be expressed in *in vitro* systems, such as reticulocyte lysate.

A further embodiment of the invention provides host cells transformed or transfected with the vectors for the replication and expression of polynucleotides of the invention. The cells

will be chosen to be compatible with the said vector and may for example be bacterial, yeast, insect or mammalian.

Polynucleotides according to the invention may also be inserted into the vectors described above in an antisense orientation in order to provide for the production of antisense RNA or ribozymes.

A still further aspect of the present invention provides a method which includes introducing the nucleic acid into a host cell. The introduction, which may (particularly for *in vitro* introduction) be generally referred to without limitation as "transformation", may employ any available technique. For eukaryotic cells, suitable techniques may include calcium phosphate transfection, DEAE-Dextran, electroporation, liposome- or polycation-mediated transfection and transduction using retrovirus or other virus, e.g. vaccinia or, for insect cells, baculovirus. For bacterial cells, suitable techniques may include calcium chloride transformation, electroporation and transfection using bacteriophage. As an alternative, direct injection of the nucleic acid could be employed.

The introduction may be followed by causing or allowing expression from the nucleic acid, e.g. by culturing host cells (which may include cells actually transformed although more likely the cells will be descendants of the transformed cells) under conditions for expression of the gene, so that the encoded polypeptide is produced. If the polypeptide is expressed coupled to an appropriate signal leader peptide it may be secreted from the cell into the culture medium. Following production by expression, a polypeptide may be isolated and/or purified from the host cell and/or culture medium, as the case may be, and subsequently used as desired, e.g. in the formulation of a composition which may include one or more additional components, such as a pharmaceutical composition which includes one or more pharmaceutically acceptable excipients, vehicles or carriers (e.g. see below).

A further aspect of the present invention provides a host cell containing nucleic acid as disclosed herein. The nucleic acid of the invention may be integrated into the genome (e.g. chromosome) of the host cell. Integration may be promoted by inclusion of sequences which promote recombination with the genome, in accordance with standard techniques. The nucleic acid may be on an extra-chromosomal vector within the cell.

Nucleic acids of the invention, particularly when in the form of a recombinant vector, may be used in methods of gene therapy. A construct capable of expressing a nucleic acid of the invention may be introduced into cells of a recipient by any suitable means, such that a polypeptide of the invention, preferably a full length or active portion polypeptide, is expressed in the cells.

The construct may be introduced in the form of naked DNA, which is taken up by some cells of animal subjects, including muscle cells of mammalians. In this aspect of the invention the construct will generally be carried by a pharmaceutically acceptable carrier alone. The construct may also formulated in a liposome particle, as described above.

Such methods of gene therapy further include the use of recombinant viral vectors such as adenoviral or retroviral vectors which comprise a construct capable of expressing a polypeptide of the invention. Such viral vectors may be delivered to the body in the form of packaged viral particles.

Constructs of the invention, however formulated and delivered, will be for use in treating conditions brought about by a defect in the *DIP* locus. The construct will comprise nucleic acid encoding the polypeptide of the invention linked to a promoter capable of expressing the gene in the target cells. The constructs may be introduced into cells of a human or non-human mammalian recipient either *in situ* or *ex-vivo* and reimplanted into the body. Where delivered *in situ*, this may

be by for example injection into target tissue(s) or in the case of liposomes, inhalation.

Gene therapy methods are widely documented in the art and may be adapted for use in the expression of a polypeptide of the invention. See for example WO95/14091 and Walther, Molecular Biotechnology, 6(3): 267-286, (1996) and Blomer, Human Molecular Genetics, Vol.5: 1397-1404, (1996), the disclosures of which are incorporated herein by reference.

F. Assays

This section (section "F") of the present application describes assays of the invention which are based on the interaction between DIP and DP-3. For the purposes of brevity, reference is made in this section to DIP or DIP protein or polypeptide, although unless stated to the contrary, this is to be taken to include the above-mentioned active portions and variants of DIP which retain the ability to interact with DP-3 or to form a homodimer.

Assays of the present invention are useful for screening putative modulator compounds which modulate the interaction between DIP and DP-3. Modulators which disrupt the interaction have putative utility compounds which can redirect the cell cycle, e.g by promoting or preventing entry into or out of a phase of the cell cycle or as promoters of apoptosis in a cell. We have found that levels of DIP mRNA are particularly prevalent in germ cells, and thus modulator compounds may be used to promote cell cycle control or modification in conditions such as teratomas or seminomas associated with abberant proliferation of germ cells. Modulators may also be used or investigated for utility in the promotion of meiosis, either *in vivo* or in *in vitro* model systems.

Definitions:

"DP protein" refers to a member of the DP family of proteins which form a transcription factor by heterodimerization with a member of the E2F family of proteins, such as E2F-1 or E2F-4.

This includes DP-1, DP-2 and DP-3 from eukaryotic sources, particularly vertebrate such as mammalian such as rodent (e.g. murine), or primate such as human. Murine DP-1 is described in, for example, WO94/10307, the disclosure of which is incorporated herein by reference. Human DP-1 and human DP-2 protein is described by Zhang and Chellappan, *Oncogene*, 1995, 10;2085-2093, the contents of which are incorporated herein by reference.

"DP-3" refers to a family member of the DP-3 family of DP proteins which are capable of forming heterodimers with members of the E2F family such as E2F-1 and E2F-4. DP-3 exists in four isoforms, α, β, γ and δ, which are described by Ormondroyd *et al*, 1995, *Oncogene*, 11;1437-1446, and in PCT/GB97/01324, the contents of which are incorporated herein by reference. The DP-3α isoform is particularly preferred for assays of the present invention. The murine DP-3α isoform is shown herein as SEQ ID NO:3.

For the purposes of the present invention, reference to DP protein includes allelic and synthetic variants of the above-described DP proteins, as well as fragments of forms of DP proteins.

A preferred group of synthetic variants include those which have at least 80%, preferably at least 90%, homology to DP-3α of SEQ ID NO:3. More preferably such variants correspond to the sequence of DP-3α of SEQ ID NO:3 but have one or more, e.g. from 1 to 10, such as from 1 to 5, substitutions, deletions or insertions of amino acids. Fragments of DP-3 and its variants are preferably at least 20, more preferably at least 50 and most preferably at least 200 amino acids in size. The DP-3 molecule of whatever isoform used will however retain the ability to physically associate *in vivo* with DIP.

For the purposes of the present invention, the precise form and structure of a DP protein or fragment thereof may be varied by

those of skill in the art, having regard to the particular assay format to be used.

Assays according to the invention may be performed *in vitro* in any format available to the person skilled in the art. The precise format of the assay of the invention may be varied by those of skill in the art using routine skill and knowledge.

For example, the interaction between DIP (as defined at the start of this section) and a protein, such as a DP-3 protein (particularly a DP-3α protein) may be studied by labeling one with a detectable label and bringing it into contact with the other which has been immobilized on a solid support. Suitable detectable labels include ^{35}S-methionine which may be incorporated into recombinantly produced DIP and/or a DP protein. The recombinantly produced DIP and/or a DP protein may also be expressed as a fusion protein containing an epitope which can be labeled with an antibody.

The protein which is immobilized on a solid support may be immobilized using an antibody against that protein bound to a solid support or via other technologies which are known *per se*. A preferred *in vitro* interaction may utilize a fusion protein including glutathione-S-transferase (GST). This may be immobilized on glutathione agarose beads. In an *in vitro* assay format of the type described above the putative modulator compound can be assayed by determining its ability to modulate the amount of labeled DIP or a DP-protein which binds to the immobilized GST-DP-protein or GST-DIP, as the case may be. This may be determined by fractionating the glutathione-agarose beads by SDS-polyacrylamide gel electrophoresis. Alternatively, the beads may be rinsed to remove unbound protein and the amount of protein which has bound can be determined by counting the amount of label present in, for example, a suitable scintillation counter.

Alternatively an antibody attached to a solid support and directed against one of DIP or DP-3 protein may be used in

place of GST to attach the molecule to the solid support. Antibodies against DIP and DP-3 proteins may be obtained in a variety of ways known as such in the art, and as discussed herein.

In an alternative mode, one of DIP and DP-protein may be labelled with a fluorescent donor moiety and the other labelled with an acceptor which is capable of reducing the emission from the donor. This allows an assay according to the invention to be conducted by fluorescence resonance energy transfer (FRET). In this mode, the fluorescence signal of the donor will be altered when DIP and the DP-protein interact. The presence to a candidate modulator compound which modulates the interaction will increase the amount of unaltered fluorescence signal of the donor.

FRET is a technique known per se in the art and thus the precise donor and acceptor molecules and the means by which they are linked to DIP and DP-protein may be accomplished by reference to the literature.

Suitable fluorescent donor moieties are those capable of transferring fluorogenic energy to another fluorogenic molecule or part of a compound and include, but are not limited to, coumarins and related dyes such as fluoresceins, rhodols and rhodamines, resorufins, cyanine dyes, bimanes, acridines, isoindoles, dansyl dyes, aminophthalic hydrazines such as luminol and isoluminol derivatives, aminophthalimides, aminonaphthalimides, aminobenzofurans, aminoquinolines, dicyanohydroquinones, and europium and terbium complexes and related compounds.

Suitable acceptors include, but are not limited to, coumarins and related fluorophores, xanthenes such as fluoresceins, rhodols and rhodamines, resorufins, cyanines, difluoroboradiazaindacenes, and phthalocyanines.

A preferred donor is fluorescein and preferred acceptors include rhodamine and carbocyanine. The isothiocyanate derivatives of these fluorescein and rhodamine, available from Aldrich Chemical Company Ltd, Gillingham, Dorset, UK, may be used to label DIP and a DP-protein. For attachment of carbocyanine, see for example Guo *et al*, J. Biol. Chem., 270; 27562-8, 1995.

Assays of the invention may also be performed *in vivo*. Such an assay may be performed in any suitable host cell, e.g a bacterial, yeast, insect or mammalian host cell. Yeast and mammalian host cells are particularly suitable.

To perform such an assay *in vivo*, constructs capable of expressing DIP and a DP-protein and a reporter gene construct may be introduced into the cells. This may be accomplished by any suitable technique, for example calcium phosphate precipitation or electroporation. The three constructs may be expressed transiently or as stable episomes, or integrated into the genome of the host cell.

In vivo assays may also take the form of two-hybrid assays wherein DIP and a DP-protein are expressed as fusion proteins, one being a fusion protein comprising a DNA binding domain (DBD), such as the yeast GAL4 binding domain, and the other being a fusion protein comprising an activation domain, such as that from GAL4 or VP16. In such a case the host cell (which again may be bacterial, yeast, insect or mammalian, particularly yeast or mammalian) will carry a reporter gene construct with a promoter comprising a DNA binding elements compatible with the DBD. The reporter gene may be a reporter gene as disclosed above. The promoters for the genes may be those discussed above.

DIP and the DP-protein and the reporter gene, may be introduced into the cell and expressed transiently or stably.

Since we have also found that the DIP gene product has a homodimerization interface, through which homodimers of DIP can form, the assay formats described above are also suitable for assaying for modulators of DIP homodimerization. These assays may be conducted as described above, by providing a second DIP polypeptide in place of the DP-protein. The assay formats described above may be used *mutatis mutandis* for this aspect of the invention.

Further assays provided by the present invention are those based upon the interaction of DIP with other cellular factors, particularly cellular factors or fragments thereof which comprise a BTB/POZ domain including such proteins mentioned herein. The assay formats described above may be used *mutatis mutandis* for this aspect of the invention.

Assays will be run with suitable controls routine to those of skill in the art.

Additional Assay Components.

The interaction between DIP and p300 which we have established may be influenced in the cell by the actions of, *inter alia*, E2F-1 and pRb. It may thus be desirable to include an E2F-1 and/or a pRb molecule in the assay of the invention. Such molecules may be included in both *in vitro* and *in vivo* assays. They may also be obtained by recombinant production, and expressed where appropriate using constructs and means analogous to those described above for the DIP and DP-3 molecules.

Cell Lines.

The assays of the invention give rise to novel cell lines useful in performing the assays. Such cell lines form a further aspect of the invention.

In a preferred aspect, cell lines of the invention will comprise:

(a) a first nucleic acid construct comprising nucleic acid encoding a DIP polypeptide operably linked to a promoter; and

(b) a second nucleic acid construct comprising nucleic acid encoding a DP-polypeptide operably linked to a promoter,

wherein at least one of said promoters, preferably both, is heterologous to the nucleic acid to which it is operably linked.

A preferred second construct encodes a DP-3 polypeptide.

The cell line may further comprise:

(c) a nucleic acid construct encoding and capable of expressing an E2F molecule, preferably E2F-1 or E2F-4.

The nucleic acid constructs may be DNA or RNA. They may be carried stably in the genome of the cell or in the form of non-integrated plasmid vectors.

Host cell lines include in particular yeast and mammalian, especially human, cell lines, including those mentioned elsewhere herein.

A putative modulator compound may be one which enhances, stabilizes or inhibits the various interactions described above, and thus reference to "modulation" includes both enhancement, stabilization or inhibition of said interactions. Putative modulator compounds include peptides based upon the domain of DIP which interacts with DP-protein, particularly peptides whose sequences are derived from the BTB/POZ domain.

The amount of modulator compounds used in assays of the invention will vary, depending upon the nature of the modulator and the format of the assay, but generally concentrations will be from the nanomolar to millimolar range, for example from 10nM to 10mM, such as from 0.1μM to 100μM.

G. DIP-targeted interactions.

The finding of DIP as a mediator of transcriptional control through its interaction with cellular factors provides novel assays for the isolation and identification of further cellular components which interact with DIP. Thus DIP may be used as a bait, for example linked to a DNA-binding domain, in a two hybrid assay system in which the baited DIP is screened against an expression library of cDNA linked to nucleic acid encoding an activation domain. This or analogous methods provides means to identify novel factors or novel interactions between DIP and known factors. Such factors or interactions will provide the basis for further assay systems for modulators of the cell cycle.

H. Transgenic Animals

In another aspect of the invention, there is provided a method for producing a transgenic non-human mammal, particularly a rodent such as a mouse, by incorporating a lesion into the locus of a DIP gene.

This may be achieved in a variety of ways. A typical strategy is to use targeted homologous recombination to replace, modify or delete the wild-type *DIP* gene in an embryonic stem (ES) cell. An targeting vector is introduced into ES cells by electroporation, lipofection or microinjection. In a few ES cells, the targeting vector pairs with the cognate chromosomal DNA sequence and transfers the desired mutation carried by the vector into the genome by homologous recombination. Screening or enrichment procedures are used to identify the transfected cells, and a transfected cell is cloned and maintained as a pure population. Next, the altered ES cells are injected into the blastocyst of a preimplantation mouse embryo or alternatively an aggregation chimera is prepared in which the ES cells are placed between two blastocysts which, with the ES cells, merge to form a single chimeric blastocyst. The chimeric blastocyst is surgically transferred into the uterus of a foster mother where the development is allowed to progress

to term. The resulting animal will be a chimera of normal and donor cells. Typically the donor cells will be from a animal with a clearly distinguishable phenotype such as skin colour, so that the chimeric progeny is easily identified. The progeny is then bred and its descendants cross-bred, giving rise to heterozygotes and homozygotes for the targeted mutation. The production of transgenic animals is described further by Capecchi, M, R., 1989, Science 244; 1288-1292; Valancius and Smithies, 1991, Mol. Cell. Biol. 11; 1402-1408; and Hasty *et al*, 1991, Nature 350; 243-246, the disclosures of which are incorporated herein by reference.

Homologous recombination in gene targeting may be used to replace the wild-type *DIP* gene with a specifically defined mutant form (e.g truncated or containing one or more substitutions).

The invention may also be used to replace the wild-type gene with a modified gene capable of expressing a wild-type or otherwise active DIP polypeptide, where the expression may be selectively blocked either permanently or temporarily. Permanent blocking may be achieved by supplying means to delete the gene in response to a signal. An example of such a means is the *cre-lox* system where phage *lox* sites are provided at either end of the transgene, or at least between a sufficient portion thereof (e.g. in two exons located either side or one or more introns). Expression of a *cre* recombinase causes excision and circularisation of the nuclei acid between the two *lox* sites. Various lines of transgenic animals, particularly mice, are currently available in the art which express *cre* recombinase in a developmentally or tissue restricted manner, see for example Tsien, Cell, Vol.87(7): 1317-1326, (1996) and Betz, Current Biology, Vol.6(10): 1307-1316 (1996). These animals may be crossed with *lox* transgenic animals of the invention to examine the function of the *DIP* gene. An alternative mechanism of control is to supply a promoter from a tetracyline resistance gene, *tet*, to the control regions of the

DIP locus such that addition of tetracyline to a cell binds to the promoter and blocks expression of the *DIP* gene.

Transgenic targeting techniques may also be used to delete the *DIP* gene. Methods of targeted gene deletion are described by Brenner *et al*, WO94/21787 (Cell Genesys), the disclosure of which is incorporated herein by reference.

Homologous recombination may also be used to produce "knock in" animals which express a polypeptide of the invention in the form of a fusion protein, fused to a detectable tag such as β-galactosidase or green fluorescent protein. Such transgenic non-human mammals may be used in methods of determining temporal and spatial expression of the *DIP* gene by monitoring the expression of the detectable tag.

A further alternative is to target control sequences responsible for expression of the *DIP* gene.

The invention extends to transgenic non-human mammals obtainable by such methods and to their progeny. Such mammals may be homozygous or heterozygous. Such mammals include mice, rodents, rabbits, sheep, goats, pigs.

Transgenic non-human mammals may be used for experimental purposes in studying the role of DIP in regulating the cell cycle and in the development of therapies designed to target the interaction of DIP with other cellular factors, particularly DP-proteins including DP-3. Such animals may be used in the study of conditions associated with abnormal regulation of germ cell proliferation, including conditions in which germ cells fail to differentiate or in which germ cell are subject to uncontrolled proliferation.

By "experimental" it is meant permissible for use in animal experimentation or testing purposes under prevailing legislation applicable to the research facility where such experimentation occurs.

The invention is illustrated by the following examples.

An isoform-specific trans activation domain in the N-terminal region of murine DP-3.

With a view towards identifying the properties of the regulated protein domains, we reasoned that the N-terminal extension that occurs in the α-isoform of DP-3 may affect its properties as a transcription factor. Thus, we investigated whether this region was transcriptionally active in mammalian cells by fusing the N-terminal 79 amino acid residues to the Gal4 DNA binding domain, referred to as G4-DP3N, and thereafter monitored the transcriptional activity of a reporter construct regulated by Gal4 binding sites.

From an activation domain-tagged library prepared from 14.5 d.p.c mouse embryos, we identified two recombinants, derived from the same gene, encoding fusion proteins capable of specifically interacting with LexA-DP-3N. Since the cDNA sequence contained in these recombinants encoded a novel protein, we have tentatively given it the designation DIP, derived from DP-Interacting Protein. The interaction between LexA-DP-3N and DIP was specific; no binding between DIP and the LexA DNA binding domain (LexDBP) or DP-3N and the Gal4 activation domain (GAD) was apparent.

DIP, a novel member of the BTB/POZ domain family of proteins.

Using the DIP cDNA, together with combined RACE and cDNA library screening (see Materials and Methods), we isolated overlapping clones that allowed us to assemble the complete sequence of murine(m) DIP. Conceptual translation of the DIP DNA sequence yielded a 524 residue polypeptide with a predicted molecular weight of 60,000. Searching the available DNA and protein sequence databases (both DNA and protein) revealed one other protein with significant similarity to mDIP. Thus, the GCL protein, the product of Drosophila melanogaster germ gene cell-less which required for proper germ cell maturation during embryogenesis, had 36% identity and 56% similarity with mDIP. The similarity was distributed throughout the protein

suggesting that mDIP is a close mammalian relative of Drosophila GCL. Furthermore, searching the available database of expressed sequence tags has identified several human sequences with more than 90% homology to mDIP at the amino acid level.

A notable feature of DIP is the presence, in the N-terminal half of the protein, of a domain previously referred to as BTB/POZ. This is found at approximately residues 90 - 198 of SEQ ID NO:2. The BTB/POZ domain is an approximately 120 residue hydrophobic-rich domain present in a variety of regulatory proteins that, generally, fall into two groups: those possessing actin binding functions and a second group with DNA binding-related properties. When compared across the BTB/POZ domain, the DIP sequence shares about 28% identity with other members of the family, although within the BTB/POZ domains of the different subgroups there is greater identity with the DNA binding group than with the actin binding group, consistent with DIP possessing a DNA related function.

Like many other BTB/POZ domain proteins, the BTB/POZ domain is located in the N-terminal region of DIP. However, usually it is the case that members of the DNA binding group have a series of zinc fingers located in the C-terminal region, whereas the actin-binding group have a glycine-rich region in the C-terminal half. In both of these respects DIP is different since it lacks both zinc fingers and a glycine-rich region, and thus may constitute another sub-group of BTB/POZ domain proteins. So far, the only other BTB/POZ domain protein that possesses similar properties is Drosophila GCL which, like DIP, has a nuclear-related function.

Further, analysis of the DIP cDNA revealed that the DIP mRNA has an extensive 3' untranslated region (UTR), of which about 1.2-1.4Kb has been sequenced. The 3' UTR contains several AREs (adenylate/uridylate-rich elements), a sequence motif known to regulate the stability of many RNAs that encode proto-oncogenes, transcription factors and cytokines (Chen and

Shyu 1995), together with two poly-adenylation signals separated by about 500 bases. Northern analysis of expression pattern of DIP RNA indicated that it is present at low levels in a wide variety of murine tissues as a transcript of about 4-4.6Kb. In certain tissues, an additional transcript was apparent at 3.5-4.2kB that may arise through the alternative utilization of the two poly-adenylation signals in the 3'UTR.

DIP is a nuclear protein in which the BTB/POZ domain is necessary for nuclear accumulation.

To assess the intracellular distribution, we expressed an epitope-tagged DIP protein in mammalian cells and thereafter immunostained. Wild-type DIP accumulated in a striking nuclear pattern that exhibited discrete speckles. Although there was some variation in the number and size of the speckles, the distribution of DIP observed here is generally consistent with other reports on the location of nuclear BTB/POZ proteins.

To gain information on the role of the domains in DIP that determine the intra-cellular distribution, we explored the properties of three derivatives of DIP. In the first, the N-terminal 46 residues were removed to create DIPΔ46, in the second the BTB/POZ domain was retained, and in the third the BTB-POZ domain was deleted to create DIP-ΔPOZ. 3μg of each plasmid was transfected into U2OS cells (see methods) and the intracellular location of the expressed protein determined by immunostaining with anti-T7 monoclonal antibody. A clear difference was apparent when the intracellular distribution of wild-type DIP was compared to DIPΔ46 specifically in the size of the nuclear speckles. For DIP-POZ, which encompasses the N-terminal region (from residue 47 up to residue 230) and contains the entire BTB/POZ domain, efficient nuclear accumulation was apparent. In contrast to wild-type and DIPΔ46, the nuclear distribution of DIP-POZ was uniform and lacked the discrete speckled appearance characteristic of the wild-type protein. In this respect, it is noteworthy that the N-terminal region retained in DIP-POZ contains a candidate nuclear localization signal (NLS) located that may be

responsible for the nuclear accumulation of DIP-POZ. DIP-ΔPOZ, which contains the C-terminal half of the protein from residue 231 to 524, failed to undergo nuclear accumulation and in many cells retained a cytoplasmic location. Overall, therefore, a number of distinct domains contribute to the intra-cellular distribution of DIP. Further, the BTB/POZ-domain is not sufficient to target the protein to the nuclear speckles characteristic of wild-type DIP, but requires additional information located in the N-terminal 46 residues and in the region C-terminal to the BTB/POZ domain.

DIP contains a dimerization domain.

Studies on several other BTB/POZ domain proteins have established that the domain functions in protein-protein interactions, allowing both the formation of homodimers or heterodimers with other BTB/POZ-domain proteins. We were interested to test the possibility that the DIP BTB/POZ-domain was capable of dimerization, an idea that we explored firstly using the yeast two-hybrid assay and secondly the equivalent mammalian cell assay. Thus, wild-type DIP or mutant derivatives were fused to either the LexA DNA binding domain or the GAD activation domain. In yeast, homodimer formation between wild-type DIP was evident, but was significantly enhanced upon deletion of the N-terminal region up to residue 231 and furthermore both bait and prey hybrid proteins containing only the C-terminal half of DIP (from residue 231) could efficiently interact. We conclude therefore that the presence of the BTB/POZ domain is not required for DIP homodimerization. Rather, these data imply that homodimerization occurs through a dedicated domain located in the C-terminal half of DIP.

A similar conclusion was reached when DIP was studied in a mammalian two-hybrid assay where DIP was fused to the Gal4 DNA binding domain, in G4-DIP, and the VP16 activation domain to DIP in VP16-DIP. Neither VP16 nor VP16-DIP was capable of stimulating the activity of Gal4 although significant stimulation occurred when G4-DIP and VP16-DIP were

co-expressed, a result consistent with the previous data derived from the yeast two-hybrid assay. Furthermore, in the absence of the BTB/POZ domain (G4-DIPΔPOZ), an interaction was still apparent with VP16-DIP, again consistent with the earlier yeast data. We conclude that the BTB/POZ domain is not required for DIP homodimerization. Rather, a C-terminal domain functions in dimerization.

DIP is a potent transcriptional repressor in mammalian cells.

We wish to determine whether DIP could activate or repress transcription. As a Gal4 fusion protein, we failed to obtain evidence that DIP could activate transcription. The ability of DIP to repress transcription was assessed by addressing the effect of G4-DIP on pSV-GAL-tk, a reported construct which contains a single Gal4 binding site between the SV40 enhancer and the herpes simplex minimal thymidine kinase promoter. Similar reporters have been employed previously to measure repression by other molecules, including G4-p107, which efficiently reduced the activity of pSV-Gal-tk. Strikingly, G4-DIP repressed transcription from pSV-GAL-tk, an effect that was specific for DNA bound DIP since repression was not evident with Gal4 DNA binding alone or when DIP was expressed without the Gal4 DNA binding domain. The level of repression caused by G4-DIP was similar to the effect of p107, the latter being a molecule with established repression properties that is believed to be involved in mediating its physiological effects. Thus, we conclude that DIP is endowed with an intrinsic ability to repress transcription in mammalian cells but, in order to do so, it needs to be targeted to the promoter context which in the G4-DIP repression assay is provided by the Gal4 DNA binding domain.

DIP interacts with DP-3 in mammalian cells.

Since we identified DIP in a yeast two-hybrid screen with the N-terminal region of DP-3 as the bait, it was necessary to test if DP-3 and DIP could associate in mammalian cells. To address this important point a number of different approaches were taken.

In the first line of investigation we studied the intracellular location of DIP and, specifically, whether co-expression of DIP with DP-3 could direct DP-3 into the nuclear speckles. Since DIP forms characteristic nuclear speckles whereas DP-3 fails to do so, we reasoned that if a similar pattern were to become apparent when DP-3 was co-expressed with DIP such a result would be compatible with the idea of a dominant influence of by DIP on DP-3, most likely through a direct interaction between both proteins.

We performed these co-localization studies with DIPΔ46 because the intense nuclear speckles caused by wild-type DIP gave a background level of non-specific co-localization. Thus, the α isoform of DP-3 was co-expressed with DIPΔ46 in a variety of mammalian cell-types. In U2OS cells, the α-isoform efficiently accumulated in nuclei, consistent with previous results. When DP-3α and DIPΔ46 were co-expressed DIPΔ46 gave a speckled distribution. However, in striking contrast to its distribution in the absence of DIP DP-3α also became concentrated in the nuclear speckles, to create an almost complete coincidence between DIP and DP-3α.

We assessed the specificity of this co-localization assay by determining if DP-3δ, which lacks the α-specific domain, was subject to a similar influence. In cells co-expressing DIPΔ46 and DP-3δ, the nuclear distribution of DIPΔ46 retained its characteristic nuclear speckles and, in the vast majority of cells, DP-3δ failed to co-localise with the DIPΔ46 speckles; very occasionally, however, DP-3δ was seen to co-localise (usually about 1 in 100 speckles). These data support the idea that the exogenous DP-3α is directed to the nuclear speckles in a DIP-dependent fashion, a process which is likely to be dependent upon a physical association between DP-3α and DIPΔ46.

To gain further support for an interaction between DP-3 and DIP, we performed two-hybrid assays in mammalian cells with an activation domain-tagged DIP, VP16-DIP, as the "prey", and either DP-3 or DP-1 fused to the Gal4 DNA binding domain

"baits". When VP16-DIP was assayed on two different DP baits that contained either the N-terminal 79 amino acid residues of DP-3α, G4-DP3N, or the equivalent region taken from DP-1, G4-DP1N, significant stimulation of activity was apparent with G4-DP3N and not with G4-DP1N. Similarly, no effect was apparent when VP16-DIP was co-expressed with the Gal4 DNA binding domain alone. These data suggest that DIP can physically interact with the N-terminal region of DP-3α in mammalian cells.

A similar assay was performed with hybrid baits in which complete DP-3α and δ sequences were fused to the Gal4 DNA binding domain. As expected, VP16-DIP enhanced activity in the presence of G4-DP3α, an effect specific for DP-3 since similar stimulation of the DNA binding domain Gal4 was not evident. Surprisingly, however, co-expression of VP16-DIP with G4-DP3δ also resulted in increased activity. Thus, although DIP can interact with the N-terminal region of DP-3α, it is likely that an additional part of DP-3 within the sequence shared by the α and δ isoforms interacts with DIP.

The region in DIP that is responsible for the common interaction with the and δ isoforms was next considered. In the mammalian two-hybrid assay, a hybrid protein in which the BTB/POZ domain was fused to the VP-16 activation domain, in VP16-POZ, enhanced the activity of both G4-DP3α and G4-DP3δ. Thus, we conclude that the DIP BTB/POZ domain is the principle region responsible for the interaction with DP-3α and δ.

To gain further support for an interaction between DIP and DP-3, we took a biochemical approach. However, a variety of procedures indicated that DIP is a highly insoluble protein, thus limiting the application of conventional approaches towards gaining biochemical evidence for an interaction between DIP and DP-3. Thus it was necessary to take an indirect approach towards establishing better evidence. To this end, we used a cell fractionation regime that enabled us to assay the

biochemical properties of DP-3 and moreover the influence of DIP.

Expression vectors for wild-type DIP or the mutant DIP proteins (ΔPOZ; Δ353-42; Δ110-144; Δ147-211) were introduced into U2OS cells, harvested and thereafter immunoblotted with an anti-DIP peptide antibody or, as a control, anti-p53 as described. The soluble fraction (S) contained soluble material extracted in lysis buffer (see methods) and the fraction indicated by I the insoluble material. The DIP proteins (with the exception of ΔPOZ) were present in the insoluble (I) material whereas in contrast the majority of endogenous p53 was present in the soluble (S) fraction.

In contrast, when DP-3α was expressed alone, it was present at about similar levels in each fraction. However, in cells expressing both DIP and DP-3α, the extraction properties of DP-3α were altered in a fashion such that DP-3α almost exclusively fractionated with DIP. Significantly, both polypeptides were present in the fraction that contained the insoluble material. A similar analysis performed with DP-3δ indicated that, like DP-3α, the polypeptide was found in each fraction. Co-expression of DIP caused subtle differences in the extraction properties of DP-3δ, but these differences were far less striking than the significant alteration that occurred for DP-3α in the presence of DIP. These data support the implications from the earlier results, namely that the altered solubility properties of DP-3α in the presence of DIP results from the physical interaction between the two proteins. The less striking change in the properties of DP-3δ in the presence of DIP is consistent with a significant role for the N-terminal extension in DP-3α in the interaction with DIP.

<u>DIP inactivates the N-terminal trans activation domain in DP-3.</u>
Since the N-terminal region of DP-3α possesses the properties of a trans activation domain, we reasoned that the physical association of DIP may modulate its transcriptional activity. We tested this possibility using G4-DP3N, which carries the

N-terminal 79 amino acid residues from the α-isoform fused to the Gal4 DBD, and efficiently activates transcription in mammalian cells. A significant reduction in the transcriptional activity of G4-DP3N occurred upon increasing DIP concentration indicating that DIP can cause a reduction in the capacity of DP-3 to activate transcription. In contrast, an effect on the basal activity of G4-DBD was not apparent.

Having established the functional consequence of the interaction between DIP and DP-3α, we wanted to identify the region in DIP that recognised the N-terminal extension. To approach this question, we used the yeast two-hybrid assay and measured the interaction between LexA-DP3N and various regions of DIP fused to GAD. A LexA-DP3N hybrid protein containing either the N-terminal 53 or 79 amino acid residues efficiently interacted with wild-type DIP. The C-terminal half of DIP was necessary for the interaction because GAD-DIP$^{47-232}$ failed to interact whereas GAD-DIP$^{231-524}$ allowed efficient activation, in contrast GAD-DIP$^{283-524}$ failed to activate. Thus, the C-terminal half of DIP efficiently interacts with the N-terminal region of DP-3α and the region in DIP between amino acid residue 231 and 284 is necessary for the interaction. Together with the previous data derived from the mammalian two-hybrid assay, we conclude that the DIP BTB/POZ domain interacts with a common region in the α and δ isoforms, and further that the N-terminal extension in interacts with a C-terminal region, outside the BTB/POZ domain in DIP.

<u>DIP is a negative regulator of cell cycle progression.</u>

We considered that the interaction between DIP and components of the E2F transcription factor may have functional consequences on cell cycle progression. To address this idea, we first assessed the effect of DIPΔ46 on the cell cycle using two different assays. In the first, expression vectors for DIPΔ46 were introduced into U2OS cells and effects on the cell cycle monitored by flow cytometry. Transfected cells were identified by introducing an expression vector for the cell surface protein CD20 and thereafter staining with the

monoclonal antibody leu16, and the cell cycle kinetics of the transfected population by propidium iodide. The effects of DIP were studied in asynchronous and synchronous populations of U2OS cells, where synchrony had been achieved with nocodazole, a treatment that blocks cell cycle progression in mitosis.

By 36h post-transfection of U2OS cells with DIPΔ46, a marked increase of about 22% in the G1 population was apparent, in contrast to the control treatment with the empty vector. Similarly, when expression vectors for DIP were introduced into U2OS cells and subsequently treated with nocodazole, an increased G1 population was apparent compared to the control treatment with many cells were arrested in the mitotic fraction. To confirm the effect of DIP, we assessed the CD20 negative population (taken from the cell population transfected with DIP and the control vector) after nocodazole treatment. Both populations were similar in their cell cycle profile, resembling that for the CD20 positive population transfected with the control vector. We conclude that exogenous DIP can promote cell cycle arrest by negatively-regulating early cell cycle progression.

As a further indication of the growth suppressing properties of DIPΔ46, we introduced DIP into U2OS cells in a vector that contained an expression cassette for the neomycin-resistance gene. After transfection, cells were grown under selection in the presence of G418 and the number of colonies determined after 14 days in culture. As expected in the absence of DIP, colony growth was apparent. In contrast, far fewer colonies were evident when DIP was expressed in the same conditions, a result consistent with the flow cytometry analysis which suggested that DIPΔ46 possesses the properties of a growth suppressor. Similar studies performed with wild-type DIP failed to demonstrate any significant effects on cell cycle kinetics.

<u>DP-3α overcomes DIP-dependent cell cycle arrest, whereas DP-3δ co-operates in apoptosis.</u>

Having determined that DIP can cause cell cycle arrest, we were interested to investigate whether this outcome could be modulated by the presence of DP-3α or δ. In U2OS cells the introduction of DP-3α or δ caused a significant increase in the proportion of cells undergoing cell cycle progression, usually the effect of DP-3α being marginally more efficient than DP-3δ in the S phase population. Thus, both DP-3α and δ are endowed with growth promoting activity.

Next, we co-expressed DP-3α or δ with DIP in U2OS cells. The presence of DP-3α overcame the growth-suppressing effect of DIP, allowing a greater proportion of cells to progress through the cell cycle. In contrast, however, co-expression of DP-3δ with DIP failed to promote cell cycle progression. Rather, there was a striking increase in the level of apoptotic cells usually apparent in a proportion of transfected cells. In contrast, the level of apoptosis in cells co-expressing DP-3α and DIP was not different from the background level (less than 1% transfected cells).

Overall, these data imply a biological relationship between DP-3 and DIP. Specifically, DP-3α and δ can alter the growth inhibitory properties of DIP, although the functional outcome differs for which of the two DP-3 is expressed. Significantly, DIP-dependent growth arrest is overcome by DP-3α. In striking contrast, the co-expression of DIP with DP-3δ stimulated apoptosis.

<u>DIP directs DP proteins and the E2F heterodimer into the nuclear speckles.</u>

Finally, since DIP showed a characteristic staining pattern when exogenously expressed, we analysed whether co-expression of DIP with DP-3 could direct DP-3 into the DIP-dependent nuclear speckles. Such a result would be compatible with the idea of a dominant influence of DIP on DP-3, most likely through a direct interaction between both proteins. These co-localization studies were performed with DIPΔ46 because this

mutant, although fully competent to bind DP-3, was able to discriminate between DP-3α and d isoforms in the biochemical experiments described earlier and gave a similar nuclear localization pattern to wild-type DIP. Thus, the a isoform of DP-3 was co-expressed with DIPΔ46 in a variety of mammalian cell-types. In U2OS cells, DP-3α efficiently accumulated in nuclei, consistent with our previous results (de la Luna *et al.* 1996), to give a nuclear diffuse staining pattern. When DP-3α and DIPΔ46 were co-expressed, in striking contrast to the distribution in the absence of DIP, DP-3α became concentrated in the nuclear speckles, resulting an almost complete coincidence between DIP and DP-3α.

We assessed the specificity of the co-localization by determining if DP-3δ, which lacks the a-specific interaction domain, was subject to a similar influence. U2OS cells were transfected with 3μg of pSV-DP3α or 3 μg of pSV-DP3δ, and each paired with either pCDNA-3 or pCMV-DIPΔ46. The intracellular distribution of the DP-3 isoforms and the DIP protein was determined with a rabbit polyclonal anti-DP-3 antibody and an anti-T7 monoclonal antibody. In cells co-expressing DIPΔ46 and DP-3δ, the nuclear distribution of DIPΔ46 retained its characteristic nuclear speckles and, in the vast majority of cells, DP-3δ failed to co-localise with the DIPΔ46 speckles; very occasionally, however, DP-3δ was seen to co-localise very faintly in some of the expressing cells, a result in agreement with the weak DP3δ/DIPΔ46 interaction apparent from the *in vitro* binding data. These data support the idea that DP-3 is directed to the nuclear speckles in a DIP-dependent fashion, a process which is likely to be caused by the physical association between the two proteins.

We next examined whether the association between DIP and DP-3α was compatible with formation of the DP-3/E2F heterodimer by assessing if an associated E2F partner could likewise be directed to the DIP-dependent nuclear speckles. U2OS cells were transfected with 2μg of pSV-DP3α and/or 2 μg of pCMV-HAE2F5 or pCMV-DIPΔ46. The intracellular distribution of the DP-3α and

E2F-5 was determined by immunofluorescence with a rabbit polyclonal antibody and an anti-HA monoclonal antibody.

Consistent with our previous studies, co-expresssion of DP-3α with E2F-5, which is predominantly cytoplasmic in the absence of a nuclear-targetting subunit such as DP-3α (Allen *et al.* 1997), caused the nuclear accumulation of E2F-5. However co-expression of DIP together with DP-3α and E2F-5 caused E2F-5 to become localised in a pattern of nuclear speckles that was co-incident with the distribution of DP-3α. Therefore, DIP can direct the E2F heterodimer into the nuclear speckles.

DIP can regulate E2F site-dependent transcription.

That DIP can interact with DP proteins prompted us to examine the possibility that this interaction reflected the ability of DIP to regulate E2F site-dependent transcription. To investigate this idea, we studied the effect of co-expressing DIP with either E2F-5 or E2F-1 together with DP-3 on the transcriptional activity of the cyclin E promoter, a cellular promoter that is known to be E2F-reponsive (Botz *et al.*, 1996; Geng *et al.*, 1996). The E2F reporter pCyclinE-luc (1μg) together with expression vectors for E2F-5 (a; 1.0 μg), E2F-1Y (b: 0.1μg), DP-3α (2.0 μg), DP-3δ (2.0 μg) or DIP (3.0 μg) were transfected into U20S cells.

In the presence of E2F-5 and either DP-3α or DP-3δ, clear activation of the cyclin E promoter was apparent. Although the co-operation between E2F and DP components was striking, there were insignificant differences between the transcriptional effects of DP-3α and DP-3δ, a result consistent with previous studies (Ormondroyd *et al.*, 1995). However, the co-expression of DIP with either of the E2F-5/DP-3 heterodimers caused a considerable reduction in the level of E2F site-dependent transcription. The transcriptional activity was usually diminished up to a level approaching 50%, and increased amounts of DIP failed to cause a greater reduction in transcriptional activity.

Similar results were apparent when the effects of DIP were assessed on the E2F-1/DP-3 heterodimer. In this experiment, we examined the effect of DIP on an E2F-1 mutant, E2F-1Y, which fails to bind to pRb (Helin *et al.*, 1993), and therefore ruled out any indirect effects that DIP may have on the activity of pRb and its interaction with E2F-1. As observed previously (Bandara *et al.*, 1993), E2F-1 could stimulate E2F site-dependent transcription in the absence of a DP partner, although the level of activity was augmented upon co-expression of a DP partner. In these conditions, the co-expression if DIP caused a reduction in the activity of the E2F-1Y/DP-3 heterodimer whilst having little apparent effect on the activity of E2F-1Y in the absence of a co-expressed DP partner. Furthermore, in a similar fashion to the effect of DIP on the E2F-5 heterodimer, DIP failed to cause the complete inactivation of the E2F-1/DP-3 heterodimer, but caused approximately a 50% reduction in the transcriptional stimulation that resulted from expression of the DP component. Overall, therefore, these results show that DIP can diminish the activity of E2F site-dependent transcription.

<u>DP proteins modulate DIP-dependent growth arrest.</u>

The data presented earlier show that DIP can cause an accumulation of G1 cells, whilst causing a reduction in both S and G2/M phase populations. An examination of the properties of the DP-3 isoforms found that, in contrast to the effects of DP-3α and DP-3δ in the transcription assays, the influence of each DP-3 isoform on cell cycle progression was quite different. To demonstrate this further, U20S cells were transfected with expression vectors for DIP (10 μg), DP-3α (10 μg) or DP-3δ (10 μg), together with the CD20 expression vector (4 μg). Backbone expression vectors were added to normalise the amounts of DNA in each transfection. Transfected cells were identified by staining with a FITC-conjugated anti-CD20 antiserum and DNA stained with propidium iodide as described in Methods. Cells transfected with CD20 and backbone expression plasmids exhibited a cell cycle profile with approximately 50%

cells in the G1 phase, 20% in the S phase and 30% in the G2/M phase of the cell cycle. The results were determined as the percentage change of cells in each phase of the cell cycle relative to CD20 expressing cells transfected with empty expression plasmids.

The most striking effect was usually apparent in the population of either S or G2/M phase cells, where a much more significant stimulation occurred upon the expression of DP-3α compared to DP-3δ. The ability of DP-3α to stimilate cell cycle progression more efficiently than DP-3δ was apparent in many different sets of data. As the only difference between the two DP-3 isoforms relates to the presence of the a isoform-specific N-terminal extension, these results thus imply a role for this region in cell cycle control. Furthermore, an analysis of E2F-5 on cell cycle progression indicated that its effects were insignificant although it could augment the activity of DP-3α when co-expressed.

We co-expressed each DP-3 isoform with DIP to determine if DIP could influence the effects of DP-3 on cell cycle progression, or vice versa and present data pertaining to the effect of DP-3α on DIP activity. The co-expression of DP-3α with DIP caused a marked effect on the cell cycle profile resulting from DIP expression. Specifically, the induction of G1 arrest was markedly reduced by co-expressing DP-3α whilst the stimulation of cell cycle progression caused by DP-3α was compromised, a conclusion particularly evident from studying the G2/M population. Under these conditions, the effect of co-expressing E2F-5 was considerable. Compared to DP-3α alone, the co-expression of DP-3α and E2F-5 with DIP further reduced the level of G1 cells and co-comittantly enhanced the S and G2/M phase population. Overall, these results argue that the cell cycle arrest caused by DIP is modulated by co-expressing DP-3.

<u>Multiple domains in DIP influence negative-growth control and the regulation of E2F activity.</u>

We sought to gather genetic evidence that DIP can influence cell cycle progression through a modulation of E2F activity and, to pursue this question, we generated a panel of DIP derivatives that were truncated at the N- or C-terminal regions, or possessed internal deletions, namely DIPΔPOZ, DIP^{46-232} $DIPD^{353-425}$, $DIPD^{232-285}$ and DIP^{1-425}. We examined the effects of these mutants on cell cycle progression and thereafter correlated their effects with the regulation of E2F site-dependent transcription, by transfecting U2OS cells with 16μg of the various vectors.

Since a striking effect of DIP was upon the size of the G1 population, we compared the G1 effect of the DIP derivatives to wild-type DIP. For example, DIPΔPOZ, which lacks the N-terminal half of DIP including the BTB/POZ domain, was compromised in G1 arrest and likewise DIP^{46-232}, which encompasses the BTB/POZ domain also showed reduced cell cycle arrest. An inspection of the DP-3-binding properties of ΔPOZ and DIP^{46-232} indicated a significant reduction in binding relative to wild-type DIP. Furthermore, and consistant with this result, neither DIP derivative had a significant effect on E2F site-dependent transcription in contrast to DIP^{1-425} which behaved in a similar fashion to wild-type DIP. These data suggest that the properties of DIP are influenced by distinct domains. They also support the importance of DP-3 as a target in DIP-dependent growth arrest.

We progressed on to analyse the properties of two additional mutant derivatives of DIP, namely $DIP^{\Delta 144-189}$. The effect of each mutant on the G1 population was quite different, as $DIP^{\Delta 232-285}$ caused a reduced cell-cycle arrest whereas, in contrast, $DIP^{\Delta 144-189}$ retained wild-type activity. Furthermore, the ability to regulate cell cycle progression correlated with the DP-3 binding activity of the two DIP mutants, as $DIP^{\Delta 232-285}$ had reduced binding activity whereas $DIP^{\Delta 144-189}$ bound DP-3 as efficiently as wild-type DIP.

Finally, we investigated the properties of $DIP^{\Delta 282-285}$ on the regulation of E2F site-dependent transcription and compared its characteristics with those of wild-type DIP. In contrast to the inactivation of E2F-dependent transcription that resulted from co-expression of wild-type DIP, $DIP^{\Delta 232-285}$ failed to significantly alter the transcriptional activity of the cyclin E promoter driven by the DP-3/E2F-5 heterodimer. These data show that there is a correlation between the ability of DIP to cause cell cycle arrest, bind to DP-3 and inactivate E2F site-dependent transcription, and are consistant with the idea that $DIP^{\Delta 232-285}$ fails to affect cell cycle progression because its DP-3 binding activity is compromised.

Cloning Human DIP.

Using the cDNA of SEQ ID NO:1 as a probe on a human cDNA library under high stringency conditions, postive clones are identified. Dideoxy sequencing is performed using primer sequences present at the vector cloning site. A sequencing reaction from the 3' end of the insert provides the nucleic acid sequence of the clone. A 135 amino acid region from the C-terminal portion of the human DIP protein is shown below as SEQ ID NO:5. This sequence is found to be 98.%% identical to the corresponding region (377-511) of SEQ ID NO:2, as measured using the BLASTN algorithm with default parameters (blastn matrix:1 -3; Gap penalties: Existence 5, Extension 2; lambda 1.37; K 0.711; H 1.31).

A single stranded DNA primer is synthesised which has the following sequence:

5' ATT GAT TTC TTG AGG CCC CAC 3'(SEQ ID NO:6)

This is an exact match of, and capable of hybridising to, the complement of nucleic acid encoding residues 2 to 8 of SEQ ID NO:5, and is used as a further sequencing primer to provide further confirmation of the coding sequence of human DIP, which is determined to encode a polypeptide greater than 90%

identical to SEQ ID NO:2, as determined by the BLASTN algorithm using default parameters.

Materials and Methods

Yeast two-hybrid screening:

Yeast two-hybrid screening was performed using as a bait the amino-terminal 79 amino acids from mouse DP-3α fused to the DNA binding domain of LexA in pLex(His) (Buck *et al.* 1995). The *Saccharomyces cerevisiae* strain CTY-10d, which carries an integrated lacZ reporter gene under the control of a LexA-responsive promoter, was transformed with the DP-3α bait and a mouse 14.5 days post-coitum embryonic cDNA library fused to Gal4 activation domain (Chevray and Nathans 1992) and double transformants plated in the appropriate selective medium. About 2 x 10^6 double transformants were screened by filter assay for the induction of β-galactosidase according to standard procedures. Two positives were obtained, and plasmids containing the prey sequences were rescued and checked by back transformation with the bait into yeast. In order to get a full length cDNA clone, an F9EC cDNA library was screened by hybridization with cDNA fragments and several clones isolated and sequenced. 5'-end sequences were obtained using RACE on mouse testis RNA (Clontech) and gene specific oligonucleotides.

Sequences of the cDNA clones were determined in both strands manually (Sequenase, Amersham) or with an ABI dye terminator cycle sequencing-ready reaction kit (Perkin Elmer) and an automated DNA sequence analyser. Sequences were assembled into contigs with the Seqman programme from the DNAStar package (DNASTAR, Inc.). Database searches and sequence comparisons were done using the following programmes provided by the National Center for Biotechnology Information: BLAST (Altschul *et al.* 1990), gapped BLAST and PSI-BLAST (Altschul *et al.* 1997).

For analysing protein-protein interactions in yeast, CTY-10d cells were transformed with various combinations of plasmids expressing LexA DNA binding domain (DBD)-tagged and Gal4 activation domain (GAD)-tagged molecules. Transformants were plated in the appropriate selective media and β-galactosidase activity determined for at least three independent colonies as described (Ormondroyd *et al.* 1995).

Plasmids:

Yeast two-hybrid assay: pGAD, pLex(His), pLex-DP3α and pLex-DP3δ have been already described (Buck *et al.* 1995; Ormondroyd *et al.* 1995). LexA-DBD derivatives were constructed by cloning the appropriate DNA segments from DP-3 genes and DIP cDNA into pLex(His). For generating the Gal4-AD derivatives, plasmid pACT-II (Clontech) was used as backbone vector.

Mammalian two-hybrid assay: pG4-DBD was constructed by cloning a HindIII-EcoRI fragment from pSG424 (Sadowski and Ptashne 1989) containing the Gal4 DNA binding domain (amino acids 1-147) into pCDNA-3 (Invitrogen). pG4DP-3N and pG4DP-1N were constucted by fusing the nucleotide sequence corresponding to the first 79 amino acid of mouse DP-1 in frame with the Gal4 DNA binding domain in pG4-DBD. pDP3α-G4DBD and pDP3δ-G4DBD are carboxy-terminal fusion proteins with the G4-DBD and were made by inserting a PCR product containing G4-DBD (amino acids 1-147) into a BamHI site of pSV-DP3α and pSV-DP3δ (de la Luna *et al.* 1996). Two DIP fragments encoding amino acids 23 to 524 and 21 to 232 were fused down-stream of VP16 activation domain in pCMV-VP16/NLS (N. Shikama and N.B. La Thangue, submitted) to generate pVP16-DIP and pVP16-POZ. Reporter vector pG5E1b-luc has been already described (Lee *et al.*, 1998)

Mammalian expression vectors: the following expression vectors have been already described: pSV-DP3α and pSV-DP3δ (de la Luna *et al.* 1996), pCMV-HAE2F5 (Allen *et al.* 1997), pCMV-βgal (Zamanian and La Thangue 1993), pCyclinE-luc (Botz *et al.*, 1996) and pSG5 (Green *et al.* 1988). To construct pCMV-DIP and

pCMV-HADIP, full length DIP was cloned into pCDNA-3 and pCMV-HA1 (Lee *et al.*, 1998), respectively. For generating DIP mutanttagged expression vectors, first the sequence for a T7 epitope was cloned into pCDNA-3 to produce N-terminus T7-tagged versions. The plasmid were used as a backbone vector for inserting a NotI-EcoRV fragment (aa 47-534), a NotI-XhoI fragment (aa 47-232) and a XhoI-EcoRV fragment (aa 231-534) from DIP full length cDNA to generate pCMV-DIPΔ46, pCMV-DIP/POZ and pCMVDIPΔPOZ, respectively. All DIP derivative mutants were made using pCMV-DIP as original plasmid and by cutting, filling in and ligataing adequate endonuclease restriction sites within the DIP sequence to generate a continuous reading frame. DIP1-425: ClaI; DIPd353-425: Asp718-ClaI; DIPd353-406: Asp718-ScaI; DIPd353-399: Asp718-HindIII; DIPd232-285: XhoI-PvuII; DIPd144-189: SspI-SphI; DIPd110-145: BglII-SspI.

Transfections:

Human osteosarcoma U2OS cells were grown in Dulbecco's modified Eagles medium (DMEM) supplemented with 5% foetal calf serum (FCS). Transfections were carried out using the calcium phosphate precipitation method. Cells were plated out 24 h before transfection at 1 x 10^5 per 6-cm dish (two-hybrid assays, immunostaining and transcription assays) or 1 x 10^6 per 10-cm dish (flow cytometric analysis), washed and refed after 16 h in the presence of the DNA precipitate and harvested and processed at a final time of 36 h post-transfection. DNA amounts were kept constant by adding pCDNA-3 or pSG5 when required. pCMV-β-galactosidase was used as an internal standard for transfections. Luciferase and β-galactosidase activities were measured in duplicate plates for each point.

For the G418 selection cells on 6-cm dishes were transfected as described above. After washing cells were trypsinized and plated at 1:10 dilution in 10-cm dishes. Next day antibiotic selection was applied starting at 500μg/ml G418 (Sigma). Cells were refed with medium with fresh antibiotic every three-four days until colonies were apparent (about three weeks). Cells

were washed with PBS, fixed with 10% formaldehyde and stained with 1% crystal violet (w/v) in 1% formaldehyde.

To arrest cells at the G2/M phase of the cell cycle, cells transfected as described above were treated with 40ng/ml nocodazole in DMSO for 12h prior to harvesting. Control cells were treated with DMSO alone.

Protein expression was checked for DIP and its derivatives by western blotting using total cell extracts. For that, transfected cells were harvested in PBS, pelleted and resuspended in 1 x SDS sample buffer. Proteins were electrophoresed in 15% SDS-polyacrylamide gels, transferred to Immobilon-P membranes (Millipore) and bands detected using ECL detection (Calbiochem). A rabbit polyclonal serum raised against a DIP specific peptide was used at a 1:500 dilution and peroxidase-conjugated anti-rabbit antibody (Amersham) was used as secondary antibody. For the solubility assay, transfected cells were harvested and resuspended in lysis buffer (50mM Tris-HCl pH8.0, 400mM KCl, 1mM EDTA, 1mM DTT, 1%NP-40 and a protease inhibitor cocktel from Boehringer Manheim). AFter being kept on ice for 30 min to allow solubilization, extracts were centrifuged at 5000 rpm for 5 min. Supernatants from this step were considered as the soluble fraction (S fraction). The insoluble material contained in the pellets (I fraction) was resuspended in SDS-sample buffer. Proteins were detected as described above using the anti-DIP polyclonal antibody and the anti-p53 monoclonal antibody DO1 (Santa Cruz).

Immunostaining:

Cells grown on coverslips were washed in PBS and treated at room temperature as follows with PBS washings after each step: fixation in 4% paraformaldehyde in PBS for 15 min, permeabilization in 1% Triton X-100 in PBS for 10 min, blocking in 5% FCS in PBS for 15 min, incubation with primary antibody in 1% FCS in PBS for 30 min and incubation with secondary antibodies in 10% FCS in PBS for 30 min. Coverslips were

mounted in Citifluor (Citifluor, Ltd), and cells photographed with an Olympus Bx60 microscope.

The following primary antibodies were used: anti-HA monoclonal antibody H11 (1:1000, BabCO), anti-T7 tag monoclonal antibody (1:10000, Novagen), anti-DIP polyclonal (1:200) and anti-DP3 rabbit polyclonal antibody (de la Luna *et al.* 1996). The secondary antibodies goat FITC-conjugated anti-rabbit and goat TRITC-conjugated anti-mouse (Southern Biotechnology Associates, Inc.) were used at a 1:200 dilution.

Flow cytometry:

DNA transfections included 4 μg of pCMVcd20 and 16 μg of the plasmid to assay. After transfection cells were detached by treatment with cell dissociation solution (Sigma). Approximately 1 x 10^6 cells were incubated at 4°C with 20 μl of the FITC-conjugated anti-cd20 antibody (Becton Dickinson) for 30 min. Cells were washed twice with phosphate-buffer saline (PBS) and fixed in 50% ethanol in PBS at 4°C for at least 1 h. Fixed cells were washed and resuspended in 50μg/ml propidium iodido containing 125u/ml Rnase A for 1 h at 4°C, washed with PBS and resuspended in 10μg/ml propidium iodide. Analysis was done on a Becton Dickinson fluorescence activated cell sorter using the FACscan software package. About 1 x 10^4 events were collected for each sample.

In vitro protein interaction:

The [^{35}S]-labelled proteins were synthesised *in vitro* using the Promega TNT kit and T7 RNA polymerase in the presence of [^{35}S]-methionine.

For immunoprecipitations, *in vitro* translated products were diluted in 200 μl of TNN buffer (50 mM Tris-HCl pH 7.4, 120 mM NaCl, 0.5% NP-40, 1 mM DTT, 1 mM PMSF and 20 μg/ml aprotinin) and incubated with 20 μl of a 50% slurry (w/v) of protein G-agarose beads preincubated with HA11 monoclonal antibody and

washed in TNN. After an incubation of 1 h at 4°C, the beads were washed four times with TNN buffer and the proteins released in SDS sample buffer and detected by SDS PAGE followed by autoradiograph.

For interactions with GST proteins, *in vitro* translated products were diluted in 200µl incubation buffer (50mM Tris pH8, 100mM NaCl, 0.5% NP40, 1mM DTT, 0.5mM EDTA, 0.2mM EGTA and 1mM PMSF). Approximately 1µg GST fusion proteins or GST protein alone purified from bacteria as described in Ormondroyd *et al.* (1995), was added in a total of 20µl glutathione agarose beads and the proteins incubated at 4°C for 5h. The beads were washed 4 times in incubation buffer and bound proteins detected by SDS PAGE followed by autoradiography. Auroradiographs were quantitated by densitometry scanning using a Biorad GS670 Imaging Densitometer.

All publications and patent applications cited in this specification are herein incorporated by reference as if each individual publication or patent application were specifically and individually indicated to be incorporated by reference.

References

Albagli, O., P.*et al* 1995. Cell Growth & Diff. 6:1193-1198.

Allen, K.E.,*et al* 1997 J. Cell Sci. 110:2819-2831.

Altschul, S.F.,*et al* 1990. J. Mol. Biol. 215:403-10.

Altschul, S.F.,*et al* 1997. Nucleic Acids Res. 25:3389-3402.

Bardwell, V.J. and R. Treisman. 1994. Genes Dev. 8:1664-677.

Bandara, L.R.,*et al* 1993. EMBO J. 13:4317-4324.

Botz, J.,*et al* 1996. Mol. Cell. Biol. 16:3401-3409.

Buck, V.,*et al* 1995. Oncogene 11:31-38.

Chen, C-Y.A. and A-B. Shyu. 1995. Trends Biochem. Sci. 20:465-470.

Chevray, P.M. and D. Nathans. 1992. Proc. Natl. Acad. Sci. 89: 5789-5793.

de la Luna, S.,*et al* 1996. J. Cell Sci. 109:2443-2452.

Geng, Y.,*et al* 1996. Oncogene 12:1173-1180.

Girling, R.,*et al* 1993. Nature 362:83-87.

Green, S.,*et al* 1988. Nucleic Acids Res. 16:369.

Helin, K.,*et al* 1993. Mol. Cell. Biol. 13:6501-6508.

La Thangue, N.B. 1994. Sci. 19:108-114.

Lam, E.W.-F. and N.B. La Thangue. 1994. Curr. Op. Cell. Biol. 6:859-866.

Lee, C-W.,*et al* 1998. Oncogene 16:2695-2710.

Nevins, J. R. 1992. Science 258:424-429.

Ormondroyd, E.,*et al* 1995. Oncogene 11:1437-1446.

Rogers, K.T.,*et al* 1996. Proc. Nat. Acad. SCI. USA 93:7594-7599.

Sadowski, I., and M. Ptashne. 1989. Nucleic Acids Res. 17:7539.

Wu, C.L.,*et al* 1995. Mol. Cell. Biol. 15:2536-2546.

Zamanian, M. and N.B. La Thangue. 1993. Mol. Biol. Cell 4:389-396.

Zhang, Y.H. and S.P. Chellappan. 1995. Oncogene 10:2085-2093.

Zollman, S., *et al* 1994. Proc. Nat. Acad. SCI. USA 91:10717-10721.

CLAIMS

1. An isolated polypeptide which comprises residues 1 to 542 of SEQ ID NO:2.

2. An isolated polypeptide having at least 70% sequence identity to SEQ ID NO:2.

3. An isolated polypeptide having at least 95% sequence identity to SEQ ID NO:2 or having at least 95% sequence identity to a portion of SEQ ID NO:2 of at least 50 contiguous amino acids.

4. An isolated polypeptide which comprises a fragment of a polypeptide as defined in claim 1 or 2, said fragment being an active fragment and/or comprising an epitope.

5. An antibody capable of binding a polypeptide according to any one of the preceding claims.

6. An antibody according to claim 5 which is a monoclonal antibody.

7. An isolated nucleic acid which encodes a polypeptide as defined in any one of claims 1 to 4, or an isolated nucleic acid complementary thereto.

8. An isolated nucleic acid which comprises a sequence having at least 70% homology to the nucleic acid sequence of SEQ ID NO:1 or nuclotides 199 to 1770 thereof, or their complements.

9. An oligonucleotide which consists essentially of from 15 to 50 contiguous nucleotides of the nucleic acid of any one of claims 7 or 8.

10. A vector which comprises a nucleic acid according to claim 7 or 8 operably linked to a promoter heterologous to said nucleic acid.

11. A host cell comprising a vector according to claim 10, the promoter of said vector being compatible with the host cell.

12. A method of detecting the presence or absence of a *DIP* gene in a nucleic acid sample, said method comprising;

(a) bringing the sample of nucleic acid into contact, under hybridizing conditions, with a polynucleotide which is selected from the group consisting of (i) SEQ ID NO:1; (ii) a variant or mutant of SEQ ID NO:1 having at least 70% homlogy to (i); and (iii) a fragment or (i) or (ii) of at least 15 nucleotides; and

(b) determining whether said polynucleotide has been able to hybridize to a homologous sequence in the sample.

13. A method according to claim 12 wherein said polynucleotide is a primer suitable for use in a polymerase chain reaction (PCR), and the determining is performed in conjuction with a second primer using PCR such that a portion of the *DIP* gene is amplified.

14. An assay for a putative modulator of DIP-DP-3 complex formation which comprises bringing into contact:

(a) a polypeptide of SEQ ID NO:2 or a variant or fragment thereof capable of binding to a DP-protein;

(b) a DP-protein or a fragment thereof which is capable of binding to a polypeptide of SEQ ID NO:2; and

(c) a putative modulator;

and measuring the degree to which said modulator is able to modulate the binding of (a) to (b).

15. An assay according to claim 14 wherein said DP-protein is a DP-3 protein.

16. An assay according to claim 15 wherein said DP-3 protein comprises the N-terminal region of DP-3α.

17. An assay according to any one of claims 14 to 17 which further includes the presence of an E2F protein.

18. An assay for a putative modulator of DIP homodimerization which assay comprises bringing two DIP polypeptides into contact with each other, under conditions suitable for dimerization to occur, in the presence of a putative modulator and measuring the degree to which dimerization is modulated.

19. An assay according to any one of claims 14 to 18 which is a two-hybrid assay.

20. A transgenic non-human mammal whose germ cells and somatic cells contain a lesion at the locus of the *DIP* gene as a result of chromosomal incorporation into the mammal genome, or into the genome of an ancestor of said mammal, of a DNA sequence which results in altered expression of the *DIP* gene.

SEQUENCE LISTING

SEQ ID NO:1 - DIP cDNA, double stranded
ORF = 199-1770

```
       1  CGGCGCGCGC TTTGTTGGGA GAAGGAGGGG GCGAGGTCTA GCGAAGCCGG

      51  TCCGCTGAGG CGCTACGGGC GGGGCTGAGG ATGGAGGTGA CTGCGTTCGC

     101  GGCGACGGTT GGGGCTGCGC GCGAGGCGGC AGCGGAGTAG GCGGTGGAGA

     151  TACGGGCATG GCGGGGCGGC CCCTCGGTGC TCCCTAGGCG CCGGAGCCAT

     201  GGGCGCTCTC AGCAGCCGGG TGCTGCGGCC CGCAGGGCGC ACAGAGCAGC

     251  CCGAACCCAC GCCCGGGGCT GGGGGCGCGG CCCGCAGGTC GGACGCCGGC

     301  GAAGATGCGG GCCACAGCTT CTGTTACTGT CCGGGCGGCC GCAAGCGCAA

     351  GCGCAGCAGC GGCACATTCT GCTACTGTCA CCCCGACTCC GAGACAGACG

     401  ACGACGAGGA CGAGGGCGAC GAGCAGCAGA GGCTGCTGAA CACGCCGCGC

     451  AGGAAAAAAT TAAAGAGCAC ATCAAAATAC ATCTACCAAA CGCTGTTTTT

     501  GAATGGTGAA AACAGTGACA TTAAGATCTG TGCTCTAGGT GAAGAGTGGA

     551  GCTTACACAA AATCTACTTA TGTCAATCTG GCTACTTTTC TAGTATGTTC

     601  AGTGGTTCTT GGAAAGAATC CAGCATGAAT ATTATTGAAC TGGAGATTCC

     651  TGACCAGAAC ATTGATATAG AAGCACTGCA GGTCGCATTT GGATCACTGT

     701  ATCGAGATGA CGTCTTAATA AAGCCCAGCA GGGTCGTTGC CATTTTGGCA

     751  GCAGCTTGCA TGCTGCAATT GGATGGTTTG ATACAGCAGT GCGGTGAGAC

     801  AATGAAGGAG ACCATCTCTG TGAGAACTGT GTGTGGCTAT TACACATCGG

     851  CAGGGACCTA TGGACTAGAC TCTGTAAAGA AAAAGTGCCT CGAGTGGCTG

     901  CTGAACAACC TCATGACTCA CCAGAGTGTG GAGCTTTTCA AAGAACTCAG

     951  TATAAACGTC ATGAAACAGC TCATTGGTTC GTCTAACTTA TTTGTGATGC

    1001  AAGTGGAGAT GGATGTATAT ACAGCTCTTA AAAAGTGGAT GTTCCTTCAG

    1051  CTGGTGCCTT CCTGGAATGG GTCTTTAAAG CAGCTTTTGA CAGAAACAGA

    1101  TGTCTGGTTT TCAAAGTGGA AAAAAGACTT TGAAGGGACG ACTTTCCTTG

    1151  AAACTGAGCA GGGAAAACCA TTTGCGCCCG TGTTCAGACA TTTAAGGCTA

    1201  CAGTACATTA TCAGTGATCT GGCTTCTGCA AGGATCATTG AGCAGGATTC

    1251  TCTGGTACCT TCAGAATGGC TGGCGGCAGT GTATAAACAG CAGTGGCTGG
```

1301 CTATGCTACG GCTGAACAA GACAGTGAAG TGGGGCCTCA AGAAATCAAT

1351 AAAGAAGAAC TTGAGGGAAA CAGCATGAGG TGTGGTCGAA AGCTTGCCAA

1401 AGATGGTGAG TACTGCTGGC GCTGGACAGG CTTCAATTTC GGCTTTAACC

1451 TCCTTGTGAC TTACACCAAT CGATACATCA TTTTCAAACG CAATACGCTG

1501 AACCAGCCAT GTAGTGGATC TGTCAGCTTA CAGCCTCGAA GGAGCATAGC

1551 ATTTAGATTG CGCTTGGCTT CTTTTGACAG TAGTGGGAAA CTCATATGCA

1601 GTAGAGCAAC TGGCTACCAA ATACTGACGC TTGAAAAAGA CCAGGAGCAA

1651 GTGGTGATGA ACTTGGACAG CAGACTTCTG ATCTTCCCTC TGTACATCTG

1701 CTGTAACTTC TTGTATATAT CACCAGAAAA AAGAACTGAG AGTAATCGTC

1751 ACCCAGAAAA CCCAGGACAC TGAGGCACTC ATCAGTGGCC AGTTTTAACT

1801 TAATGACCTA CTGCGTTCAC GTCCAAGGTG ACTAACAGTG ACCGGCCTTA

1851 TGAACTGTGG GACCCTGGAG ATGTCCTCAC CCTCATTACA TTTCTATGCA

1901 CATATGAAAA AGTTTTTTAA AACTGAGAAA GCATCTGTCA AACCATGTTA

1951 AAAGGATATC AACTCTTGCT TTAATTTAGT AGCAGTAAAA ATTGCTGTAG

2001 GTAAATTTCT CATTTCTTTG CAACAAGATA TAGATTTAAT TTTGAGCTTG

2051 AATTTCGATC CTATCTAATG TTAGTGAGTT TACTCATCTG TAAATGTGTT

2101 CCTGTTTTGT TAAGAGAATG CTAAGGACGG GAGTTTAAGT GGCCAATCAT

2151 AAATGCTCTT TCAATTGGTG CCTTTAAAGC TGTCCTTTCA TTTGTTCAGC

2201 CACCTGGTGA AGTCTTGTGT TCAGTTGAGA TCCTTTCCAG AATCTTCTTC

2251 GTTTTAGCCA TCAGATCCCT TGTCATGGTT TGGGAGATTA AGCTTAAGAG

2301 TTGGTTTTGT TTCGTTTTTC TTTTTTCTTT TTTTATTCCT AAATATAAGC

2351 ATTTCCATAT TTTTTTAAAC TTTAACTTTC CTACAGTTTT AAATGACAAC

2401 AAGAATCAGC TTGAGGACTT GAAATAAACT TGAATAAGGG AAATCAGTTT

2451 GATTCTGAGG ATATTCCTTC GCCTTACGTG CATGGTCTTG TCTGGCATTC

2501 TGTGTACTCT TATGTTATGA ATTATTTATG TACAGATACA TCGTCACACA

2551 TTTTCATATT CTCTGATAGT TGGAATTTAT GTCAAATTTT AGTTGGGAAT

2601 CTTGTTTCCT ATTAAGCTCA GGACTTTATA CCCTCTGAAC TGAGTGCCTT

2651 TGAGTCACAT ACATAAGTAG ATTTCATGGT GAATGTACAG GGGGAGCTGG

2701 GAACAAAACA GTCGTGTTAG AACAACTTAC CTTGTCAAGA ATAAGATGTA

2751 TGCTCATCCC GATCTGTATG TGTTGGTCCC TTGTGTTTAT TGCTATTGAA

2801 AATAGCCTGG ATCATTGTCC TCCAGTCAGA GTAAGTGTTT TTGTGGTTGA

2851 TGTATGTTTT TAGTATTTTT TTAAATTGGA ATCATTTTTA TGTATCTGTT

2901 CAAATAATAA ATGCTCATTT GGAAAAGGAA AAAAAAAAAA AAAAAAAAAA

2951 AAAAAAAA

SEQ ID NO:2:DIP amino acid sequence: Length: 524

1 MGALSSRVLR PAGRTEQPEP TPGAGGAARR SDAGEDAGHS FCYCPGGRKR

51 KRSSGTFCYC HPDSETDDDE DEGDEQQRLL NTPRRKKLKS TSKYIYQTLF

101 LNGENSDIKI CALGEEWSLH KIYLCQSGYF SSMFSGSWKE SSMNIIELEI

151 PDQNIDIEAL QVAFGSLYRD DVLIKPSRVV AILAAACMLQ LDGLIQQCGE

201 TMKETISVRT VCGYYTSAGT YGLDSVKKKC LEWLLNNLMT HQSVELFKEL

251 SINVMKQLIG SSNLFVMQVE MDVYTALKKW MFLQLVPSWN GSLKQLLTET

301 DVWFSKWKKD FEGTTFLETE QGKPFAPVFR HLRLQYIISD LASARIIEQD

351 SLVPSEWLAA VYKQQWLAML RAEQDSEVGP QEINKEELEG NSMRCGRKLA

401 KDGEYCWRWT GFNFGFNLLV TYTNRYIIFK RNTLNQPCSG SVSLQPRRSI

451 AFRLRLASFD SSGKLICSRA TGYQILTLEK DQEQVVMNLD SRLLIFPLYI

501 CCNFLYISPE KRTESNRHPE NPGH

SEQ ID NO:3 - DP-3, α-isoform.

ATG ACG GCA AAA AAT GTT GGT TTG CCA TCC ACA AAT GCA GAG CTG AGG
48
Met Thr Ala Lys Asn Val Gly Leu Pro Ser Thr Asn Ala Glu Leu Arg
1 5 10 15

GGC TTT ATA GAT CAG AAT TTC AGT CCA ACG AAA GGT AAC ATT TCA CTT
96
Gly Phe Ile Asp Gln Asn Phe Ser Pro Thr Lys Gly Asn Ile Ser Leu
20 25 30

GTT GCC TTT CCA GTT TCA AGC ACC AAC TCA CCA ACA AAG ATT TTA CCG
144
Val Ala Phe Pro Val Ser Ser Thr Asn Ser Pro Thr Lys Ile Leu Pro
35 40 45

AAA ACC TTA GGG CCA ATA AAT GTG AAT GTT GGA CCC CAA ATG ATT ATA
192

Lys Thr Leu Gly Pro Ile Asn Val Asn Val Gly Pro Gln Met Ile Ile
50 55 60

AGC ACA CCG CAG AGA ATT GCC AAT TCA GGA AGT GTT CTG ATT GGG AAT
240
Ser Thr Pro Gln Arg Ile Ala Asn Ser Gly Ser Val Leu Ile Gly Asn
65 70 75 80

CCA TAT ACC CCT GCA CCC GCA ATG GTC ACT CAG ACT CAC ATA GCT GAG
288
Pro Tyr Thr Pro Ala Pro Ala Met Val Thr Gln Thr His Ile Ala Glu
85 90 95

GCT GCT GGC TGG GTT CCC AGT GAT AGA AAA CGA GCT AGA GAA TTT ATA
336
Ala Ala Gly Trp Val Pro Ser Asp Arg Lys Arg Ala Arg Glu Phe Ile
100 105 110

GAC TCT GAT TTT TCA GAA AGT AAA CGA AGC AAA AAA GGA GAT AAA AAT
384
Asp Ser Asp Phe Ser Glu Ser Lys Arg Ser Lys Lys Gly Asp Lys Asn
115 120 125

GGG AAA GGC TTG AGA CAT TTT TCA ATG AAG GTG TGT GAG AAA GTT CAG
432
Gly Lys Gly Leu Arg His Phe Ser Met Lys Val Cys Glu Lys Val Gln
130 135 140

CGG AAA GGC ACA ACT TCA TAC AAT GAG GTA GCT GAT GAG CTG GTA TCT
480
Arg Lys Gly Thr Thr Ser Tyr Asn Glu Val Ala Asp Glu Leu Val Ser
145 150 155 160

GAG TTT ACC AAC TCA AAT AAC CAT CTG GCA GCT GAT TCG GCT TAT GAT
528
Glu Phe Thr Asn Ser Asn Asn His Leu Ala Ala Asp Ser Ala Tyr Asp
165 170 175

CAG GAG AAC ATT AGA CGA AGA GTT TAT GAT GCT TTA AAT GTA CTA ATG
576
Gln Glu Asn Ile Arg Arg Arg Val Tyr Asp Ala Leu Asn Val Leu Met
180 185 190

GCG ATG AAC ATA ATT TCA AAG GAA AAA AAA GAA ATC AAG TGG ATT GGC
624
Ala Met Asn Ile Ile Ser Lys Glu Lys Lys Glu Ile Lys Trp Ile Gly
195 200 205

CTG CCT ACC AAT TCT GCT CAG GAA TGC CAG AAC CTG GAA ATC GAG AAG
672
Leu Pro Thr Asn Ser Ala Gln Glu Cys Gln Asn Leu Glu Ile Glu Lys
210 215 220

CAG AGG CGG ATA GAA CGG ATA AAG CAG AAG CGA GCC CAG CTA CAA GAA
720
Gln Arg Arg Ile Glu Arg Ile Lys Gln Lys Arg Ala Gln Leu Gln Glu

```
225                 230                 235                 240

CTT CTC CTT CAG CAA ATT GCT TTT AAA AAC CTG GTA CAG AGA AAT CGA
768
Leu Leu Leu Gln Gln Ile Ala Phe Lys Asn Leu Val Gln Arg Asn Arg
                245                 250                 255

CAA AAT GAA CAA CAA AAC CAG GGC CCT CCA GCT GTG AAT TCC ACC ATT
816
Gln Asn Glu Gln Gln Asn Gln Gly Pro Pro Ala Val Asn Ser Thr Ile
            260                 265                 270

CAG CTG CCA TTT ATA ATC ATT AAT ACA AGC AGG AAA ACA GTC ATA GAC
864
Gln Leu Pro Phe Ile Ile Ile Asn Thr Ser Arg Lys Thr Val Ile Asp
        275                 280                 285

TGC AGC ATC TCC AGT GAC AAA TTT GAA TAC CTT TTT AAT TTT GAT AAC
912
Cys Ser Ile Ser Ser Asp Lys Phe Glu Tyr Leu Phe Asn Phe Asp Asn
    290                 295                 300

ACC TTT GAG ATC CAC GAC GAC ATA GAG GTA CTG AAG CGG ATG GGA ATG
960
Thr Phe Glu Ile His Asp Asp Ile Glu Val Leu Lys Arg Met Gly Met
305                 310                 315                 320

TCC TTT GGT CTG GAG TCA GGC AAA TGC TCT CTG GAG GAT CTG AAA ATC
1008
Ser Phe Gly Leu Glu Ser Gly Lys Cys Ser Leu Glu Asp Leu Lys Ile
                325                 330                 335

GCA AGA TCC CTG GTT CCA AAA GCT TTA GAA GGC TAT ATT ACA GAT ATC
1056
Ala Arg Ser Leu Val Pro Lys Ala Leu Glu Gly Tyr Ile Thr Asp Ile
            340                 345                 350

TCC ACA GGA CCT TCT TGG TTA AAT CAG GGA CTA CTT TTG AAC TCT ACC
1104
Ser Thr Gly Pro Ser Trp Leu Asn Gln Gly Leu Leu Leu Asn Ser Thr
        355                 360                 365

CAA TCA GTT TCA AAT TTA GAC CCG ACC ACC GGT GCC ACT GTA CCC CAA
1152
Gln Ser Val Ser Asn Leu Asp Pro Thr Thr Gly Ala Thr Val Pro Gln
    370                 375                 380

TCA AGT GTA AAC CAA GGG TTG TGC TTG GAT GCT GAA GTG GCC TTA GCA
1200
Ser Ser Val Asn Gln Gly Leu Cys Leu Asp Ala Glu Val Ala Leu Ala
385                 390                 395                 400

ACT GGG CAG CTC CCT GCC TCA AAC AGT CAC CAG TCC AGC AGT GCA GCC
1248
Thr Gly Gln Leu Pro Ala Ser Asn Ser His Gln Ser Ser Ser Ala Ala
                405                 410                 415
```

```
TCT CAC TTC TCG GAG TCC CGC GGC GAG ACC CCC TGT TCA TTC AAC GAT
1296
Ser His Phe Ser Glu Ser Arg Gly Glu Thr Pro Cys Ser Phe Asn Asp
            420                 425                 430

GAA GAT GAG GAA GAT GAA GAG GAG GAT CCC TCC TCC CCA GAA
1338
Glu Asp Glu Glu Asp Glu Glu Glu Asp Pro Ser Ser Pro Glu
        435                 440                 445

TAAAGACAGG AGAGAACTCA TGTTTTAAAA AAAAAAAAAA ACTCGAG
1385
```

SEQ ID NO.4

Translation of SEQ ID NO.3

SEQ ID NO:5

Human DIP - C terminal region

```
        1 EVGPQEINKE ELEGNSMRCG RKLAKDGEYC WRWTGFNFGF DLLVTYTNRY
IIFKRNTLNQ
       61 PCSGSVSLQP RRSIAFRLRL ASFDSSGKLI CSRTTGYQIL TLEKDQEQVV
MNLDSRLLIF
      121 PLYICCNFLY ISPEK
```

SEQ ID NO.6

Primer - single stranded DNA
Complementary to human DIP coding sequence
5' ATT GAT TTC TTG AGG CCC CAC 3'

Although the foregoing invention has been described in some detail by way of illustration and example for the purposes of clarity and understanding, it will be readily apparent to those of skill in the art that certain changes and modifications may be made thereto without departing from the spirit or scope of the appended claims.

INTERNATIONAL SEARCH REPORT

Internati Application No
PCT/GB 98/03485

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 C12N15/12 C07K14/47 C07K16/18 C12N15/85 C12N5/10
C12Q1/68 G01N33/68 A01K67/027

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 C07K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	ADAMS ET AL.: "Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence" EMBL NUCLEOTIDE DATABASE,18 April 1997, XP002100996 AC AA348897 see the whole document & NATURE, vol. 377, 1995, pages 3-174, ---	3,7,9
X	MARRA ET AL.: "The WashU-HHMI Mouse EST Project" EMBL NUCLEOTIDE DATABASE,25 June 1997, XP002100942 AC AA386864 see the whole document ---	7,9
	-/--	

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
23 April 1999	12/05/1999
Name and mailing address of the ISA European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Authorized officer Ceder, O

3

Form PCT/ISA/210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

Internat. Application No
PCT/GB 98/03485

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	MARRA ET AL.: "The WashU-HHMI Mouse EST Project" EMBL NUCLEOTIDE DATABASE,31 January 1997, XP002100943 AC AA204129 see the whole document ---	7,9
X	MARRA ET AL.: "The WashU-HHMI Mouse EST Project" EMBL NUCLEOTIDE DATABASE,13 June 1997, XP002100944 Ac AA462393 see the whole document ---	7,9
A	CHOUBEY D ET AL: "Inhibition of E2F-mediated transcription by p202." THE EMBO JOURNAL, vol. 15, no. 20, 15 October 1996, pages 5668-5678, XP002100945 see abstract see page 5669, right-hand column, line 1 - line 15 see page 5671, left-hand column ---	10,11, 14,19
A	WO 97 02354 A (MEDICAL RES COUNCIL ;LATHANGUE NICHOLAS BARRIE (GB)) 23 January 1997 -----	

INTERNATIONAL SEARCH REPORT

Information on patent family members

Internat' ' Application No
PCT/GB 98/03485

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9702354 A	23-01-1997	AU 6237396 A	05-02-1997
		CA 2224940 A	23-01-1997
		EP 0835319 A	15-04-1998